(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
21 March 2002 (21.03.2002)

PCT

(10) International Publication Number
WO 02/22888 A1

(51) International Patent Classification[7]: C12Q 1/68, C07H 21/02, 21/04, C12N 15/00

(21) International Application Number: PCT/US01/29332

(22) International Filing Date: 17 September 2001 (17.09.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/232,755 15 September 2000 (15.09.2000) US

(71) Applicant (*for all designated States except US*): **GENAISSANCE PHARMACEUTICALS, INC.** [US/US]; Five Science Park, New Haven, CT 06511 (US).

(72) Inventors; and
(75) Inventors/Applicants (*for US only*): **ANASTASIO, Alison, E.** [US/US]; 367 Alden Avenue #8B, New Haven, CT 06515 (US). **CHEW, Anne** [US/US]; 1477 Beacon Street, #64, Brookline, MA 02446 (US). **HAN, Jin-Hua** [US/US]; 569B Prospect Street, New Haven, CT 06511 (US). **LEE, Helen, H.** [US/US]; 26 Rivendell Drive, Shelton, CT 06484 (US).

(74) Agents: **SHANER, Sandra, L.** et al.; Genaissance Pharmaceuticals, Inc., Five Science Park, New Haven, CT 06511 (US).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) **Title:** HAPLOTYPES OF THE SCYA8 GENE

(57) **Abstract:** Novel genetic variants of the Small Inducible Cytokine Subfamily A (Cys-Cys), Member 8 (Monocyte Chemotactic Protein 2) (SCYA8) gene are described. Various genotypes, haplotypes, and haplotype pairs that exist in the general United States population are disclosed for the SCYA8 gene. Compositions and methods for haplotyping and/or genotyping the SCYA8 gene in an individual are also disclosed. Polynucleotides defined by the haplotypes disclosed herein are also described.

WO 02/22888 A1

HAPLOTYPES OF THE SCYA8 GENE

RELATED APPLICATIONS

This application claims the benefit of U.S. Provisional Application Serial No. 60/232,755 filed September 15, 2000.

FIELD OF THE INVENTION

This invention relates to variation in genes that encode pharmaceutically-important proteins. In particular, this invention provides genetic variants of the human small inducible cytokine subfamily A (Cys-Cys), member 8 (monocyte chemotactic protein 2) (SCYA8) gene and methods for identifying which variant(s) of this gene is/are possessed by an individual.

BACKGROUND OF THE INVENTION

Current methods for identifying pharmaceuticals to treat disease often start by identifying, cloning, and expressing an important target protein related to the disease. A determination of whether an agonist or antagonist is needed to produce an effect that may benefit a patient with the disease is then made. Then, vast numbers of compounds are screened against the target protein to find new potential drugs. The desired outcome of this process is a lead compound that is specific for the target, thereby reducing the incidence of the undesired side effects usually caused by activity at non-intended targets. The lead compound identified in this screening process then undergoes further *in vitro* and *in vivo* testing to determine its absorption, disposition, metabolism and toxicological profiles. Typically, this testing involves use of cell lines and animal models with limited, if any, genetic diversity.

What this approach fails to consider, however, is that natural genetic variability exists between individuals in any and every population with respect to pharmaceutically-important proteins, including the protein targets of candidate drugs, the enzymes that metabolize these drugs and the proteins whose activity is modulated by such drug targets. Subtle alteration(s) in the primary nucleotide sequence of a gene encoding a pharmaceutically-important protein may be manifested as significant variation in expression, structure and/or function of the protein. Such alterations may explain the relatively high degree of uncertainty inherent in the treatment of individuals with a drug whose design is based upon a single representative example of the target or enzyme(s) involved in metabolizing the drug. For example, it is well-established that some drugs frequently have lower efficacy in some individuals than others, which means such individuals and their physicians must weigh the possible benefit of a larger dosage against a greater risk of side effects. Also, there is significant variation in how well people metabolize drugs and other exogenous chemicals, resulting in substantial interindividual variation in the toxicity and/or efficacy of such exogenous substances (Evans et al., 1999, *Science* 286:487-491). This variability in efficacy or toxicity of a drug in genetically-diverse patients makes many drugs ineffective or even dangerous in certain groups of the population, leading to the failure of such drugs in

clinical trials or their early withdrawal from the market even though they could be highly beneficial for other groups in the population. This problem significantly increases the time and cost of drug discovery and development, which is a matter of great public concern.

It is well-recognized by pharmaceutical scientists that considering the impact of the genetic variability of pharmaceutically-important proteins in the early phases of drug discovery and development is likely to reduce the failure rate of candidate and approved drugs (Marshall A 1997 *Nature Biotech* **15**:1249-52; Kleyn PW et al. 1998 *Science* **281**: 1820-21; Kola I 1999 *Curr Opin Biotech* **10**:589-92; Hill AVS et al. 1999 in *Evolution in Health and Disease* Stearns SS (Ed.) Oxford University Press, New York, pp 62-76; Meyer U.A. 1999 in *Evolution in Health and Disease* Stearns SS (Ed.) Oxford University Press, New York, pp 41-49; Kalow W et al. 1999 *Clin. Pharm. Therap.* **66**:445-7; Marshall, E 1999 *Science* **284**:406-7; Judson R et al. 2000 *Pharmacogenomics* **1**:1-12; Roses AD 2000 *Nature* **405**:857-65). However, in practice this has been difficult to do, in large part because of the time and cost required for discovering the amount of genetic variation that exists in the population (Chakravarti A 1998 *Nature Genet* **19**:216-7; Wang DG et al 1998 *Science* **280**:1077-82; Chakravarti A 1999 *Nat Genet* **21**:56-60 (suppl); Stephens JC 1999 *Mol. Diagnosis* **4**:309-317; Kwok PY and Gu S 1999 *Mol. Med. Today* **5**:538-43; Davidson S 2000 *Nature Biotech* **18**:1134-5).

The standard for measuring genetic variation among individuals is the haplotype, which is the ordered combination of polymorphisms in the sequence of each form of a gene that exists in the population. Because haplotypes represent the variation across each form of a gene, they provide a more accurate and reliable measurement of genetic variation than individual polymorphisms. For example, while specific variations in gene sequences have been associated with a particular phenotype such as disease susceptibility (Roses AD *supra*; Ulbrecht M et al. 2000 *Am J Respir Crit Care Med* **161**: 469-74) and drug response (Wolfe CR et al. 2000 *BMJ* **320**:987-90; Dahl BS 1997 *Acta Psychiatr Scand* **96** (Suppl 391): 14-21), in many other cases an individual polymorphism may be found in a variety of genomic backgrounds, i.e., different haplotypes, and therefore shows no definitive coupling between the polymorphism and the causative site for the phenotype (Clark AG et al. 1998 *Am J Hum Genet* **63**:595-612; Ulbrecht M et al. 2000 *supra*; Drysdale et al. 2000 *PNAS* **97**:10483-10488). Thus, there is an unmet need in the pharmaceutical industry for information on what haplotypes exist in the population for pharmaceutically-important genes. Such haplotype information would be useful in improving the efficiency and output of several steps in the drug discovery and development process, including target validation, identifying lead compounds, and early phase clinical trials (Marshall et al., *supra*).

One pharmaceutically-important gene for the treatment of inflammatory diseases and HIV is the small inducible cytokine subfamily A (Cys-Cys), member 8 (monocyte chemotactic protein 2) (SCYA8) gene or its encoded product. SCYA8 mRNA is predominantly expressed in the small intestine, peripheral blood, heart, placenta, lung, skeletal muscle, ovary, colon, spinal cord, pancreas, and thymus (Van Coillie et al. *Genomics* 1997 Mar 1;40(2):323-31). SCYA8, also known as MCP2,

belongs to the intercrine beta family, also known as the chemokine CC gene family and to the the monocyte chemotactic protein (MCP) subfamily of chemokines. SCYA8 is chemotactic for and activates a wide variety of inflammatory cells. SCYA8 is inducible by proinflammatory cytokines in mononuclear cells and fibroblasts. Members of the MCP subfamily of chemokines recruit leukocytes to sites of inflammation and may contribute to tumor-associated leukocyte infiltration and to the antiviral state against HIV infection. In addition, this group of chemokines is believed to play a fundamental role in the development of allergic airway diseases such as asthma (Pype et al. *Am J Respir Cell Mol Biol* 1999 Oct;21(4):528-36). SCYA8 exhibits a broader spectrum of targeted cells than other members of the MCP subfamily. Therefore, SCYA8 may play an important role in recruiting/activating immune cells at inflammatory and neoplastic foci (Gong et al. *J Biol Chem* 1997 May 2;272(18):11682-5).

The small inducible cytokine subfamily A (Cys-Cys), member 8 (monocyte chemotactic protein 2) gene is located on chromosome 17 and contains 3 exons that encode a 109 amino acid protein. A reference sequence for the SCYA8 gene is shown in Figure 1(Contig No. 1097793; SEQ ID NO: 1). Reference sequences for the coding sequence (Contig No. NM_005623.1) and protein are shown in Figures 2 (SEQ ID NO: 2) and 3 (SEQ ID NO: 3), respectively.

Because of the potential for variation in the SCYA8 gene to affect the expression and function of the encoded protein, it would be useful to know whether polymorphisms exist in the SCYA8 gene, as well as how such polymorphisms are combined in different copies of the gene. Such information could be applied for studying the biological function of SCYA8 as well as in identifying drugs targeting this protein for the treatment of disorders related to its abnormal expression or function.

SUMMARY OF THE INVENTION

Accordingly, the inventors herein have discovered 12 novel polymorphic sites in the SCYA8 gene. These polymorphic sites (PS) correspond to the following nucleotide positions in Figure 1: 9599 (PS1), 9781 (PS2), 10001 (PS3), 10053 (PS4), 10061 (PS5), 10096 (PS6), 10119 (PS7), 10903 (PS8), 10948 (PS9), 11032 (PS10), 11481 (PS11) and 11496 (PS12) in 1097793. The polymorphisms at these sites are cytosine or thymine at PS1, guanine or adenine at PS2, thymine or cytosine at PS3, adenine or cytosine at PS4, guanine or cytosine at PS5, cytosine or thymine at PS6, adenine or guanine at PS7, cytosine or guanine at PS8, cytosine or thymine at PS9, cytosine or adenine at PS10, adenine or cytosine at PS11 and guanine or thymine at PS12. In addition, the inventors have determined the identity of the alleles at these sites in a human reference population of 79 unrelated individuals self-identified as belonging to one of four major population groups: African descent, Asian, Caucasian and Hispanic/Latino. From this information, the inventors deduced a set of haplotypes and haplotype pairs for PS1-PS12 in the SCYA8 gene, which are shown below in Tables 5 and 4, respectively. Each of these SCYA8 haplotypes constitutes a code that defines the variant nucleotides that exist in the human population at this set of polymorphic sites in the SCYA8 gene. Thus each SCYA8 haplotype also

represents a naturally-occurring isoform (also referred to herein as an "isogene") of the SCYA8 gene. The frequency of each haplotype and haplotype pair within the total reference population and within each of the four major population groups included in the reference population was also determined.

Thus, in one embodiment, the invention provides a method, composition and kit for genotyping the SCYA8 gene in an individual. The genotyping method comprises identifying the nucleotide pair that is present at one or more polymorphic sites selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11 and PS12 in both copies of the SCYA8 gene from the individual. A genotyping composition of the invention comprises an oligonucleotide probe or primer which is designed to specifically hybridize to a target region containing, or adjacent to, one of these novel SCYA8 polymorphic sites. A genotyping kit of the invention comprises a set of oligonucleotides designed to genotype each of these novel SCYA8 polymorphic sites. The genotyping method, composition, and kit are useful in determining whether an individual has one of the haplotypes in Table 5 below or has one of the haplotype pairs in Table 4 below.

The invention also provides a method for haplotyping the SCYA8 gene in an individual. In one embodiment, the haplotyping method comprises determining, for one copy of the SCYA8 gene, the identity of the nucleotide at one or more polymorphic sites selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11 and PS12. In another embodiment, the haplotyping method comprises determining whether one copy of the individual's SCYA8 gene is defined by one of the SCYA8 haplotypes shown in Table 5, below, or a sub-haplotype thereof. In a preferred embodiment, the haplotyping method comprises determining whether both copies of the individual's SCYA8 gene are defined by one of the SCYA8 haplotype pairs shown in Table 4 below, or a sub-haplotype pair thereof. Establishing the SCYA8 haplotype or haplotype pair of an individual is useful for improving the efficiency and reliability of several steps in the discovery and development of drugs for treating diseases associated with SCYA8 activity, e.g., inflammatory diseases and HIV.

For example, the haplotyping method can be used by the pharmaceutical research scientist to validate SCYA8 as a candidate target for treating a specific condition or disease predicted to be associated with SCYA8 activity. Determining for a particular population the frequency of one or more of the individual SCYA8 haplotypes or haplotype pairs described herein will facilitate a decision on whether to pursue SCYA8 as a target for treating the specific disease of interest. In particular, if variable SCYA8 activity is associated with the disease, then one or more SCYA8 haplotypes or haplotype pairs will be found at a higher frequency in disease cohorts than in appropriately genetically matched controls. Conversely, if each of the observed SCYA8 haplotypes are of similar frequencies in the disease and control groups, then it may be inferred that variable SCYA8 activity has little, if any, involvement with that disease. In either case, the pharmaceutical research scientist can, without *a priori* knowledge as to the phenotypic effect of any SCYA8 haplotype or haplotype pair, apply the information derived from detecting SCYA8 haplotypes in an individual to decide whether modulating SCYA8 activity would be useful in treating the disease.

The claimed invention is also useful in screening for compounds targeting SCYA8 to treat a specific condition or disease predicted to be associated with SCYA8 activity. For example, detecting which of the SCYA8 haplotypes or haplotype pairs disclosed herein are present in individual members of a population with the specific disease of interest enables the pharmaceutical scientist to screen for a compound(s) that displays the highest desired agonist or antagonist activity for each of the SCYA8 isoforms present in the disease population, or for only the most frequent SCYA8 isoforms present in the disease population. Thus, without requiring any *a priori* knowledge of the phenotypic effect of any particular SCYA8 haplotype or haplotype pair, the claimed haplotyping method provides the scientist with a tool to identify lead compounds that are more likely to show efficacy in clinical trials.

Haplotyping the SCYA8 gene in an individual is also useful in the design of clinical trials of candidate drugs for treating a specific condition or disease predicted to be associated with SCYA8 activity. For example, instead of randomly assigning patients with the disease of interest to the treatment or control group as is typically done now, determining which of the SCYA8 haplotype(s) disclosed herein are present in individual patients enables the pharmaceutical scientist to distribute SCYA8 haplotypes and/or haplotype pairs evenly to treatment and control groups, thereby reducing the potential for bias in the results that could be introduced by a larger frequency of a SCYA8 haplotype or haplotype pair that is associated with response to the drug being studied in the trial, even if this association was previously unknown. Thus, by practicing the claimed invention, the scientist can more confidently rely on the information learned from the trial, without first determining the phenotypic effect of any SCYA8 haplotype or haplotype pair.

In another embodiment, the invention provides a method for identifying an association between a trait and a SCYA8 genotype, haplotype, or haplotype pair for one or more of the novel polymorphic sites described herein. The method comprises comparing the frequency of the SCYA8 genotype, haplotype, or haplotype pair in a population exhibiting the trait with the frequency of the SCYA8 genotype or haplotype in a reference population. A higher frequency of the SCYA8 genotype, haplotype, or haplotype pair in the trait population than in the reference population indicates the trait is associated with the SCYA8 genotype, haplotype, or haplotype pair. In preferred embodiments, the trait is susceptibility to a disease, severity of a disease, the staging of a disease or response to a drug. In a particularly preferred embodiment, the SCYA8 haplotype is selected from the haplotypes shown in Table 5, or a sub-haplotype thereof. Such methods have applicability in developing diagnostic tests and therapeutic treatments for inflammatory diseases and HIV.

In yet another embodiment, the invention provides an isolated polynucleotide comprising a nucleotide sequence which is a polymorphic variant of a reference sequence for the SCYA8 gene or a fragment thereof. The reference sequence comprises the contiguous sequences shown in Figure 1 and the polymorphic variant comprises at least one polymorphism selected from the group consisting of thymine at PS1, adenine at PS2, cytosine at PS3, cytosine at PS4, cytosine at PS5, thymine at PS6, guanine at PS7, guanine at PS8, thymine at PS9, adenine at PS10, cytosine at PS11 and thymine at

PS12.

A particularly preferred polymorphic variant is an isogene of the SCYA8 gene. A SCYA8 isogene of the invention comprises cytosine or thymine at PS1, guanine or adenine at PS2, thymine or cytosine at PS3, adenine or cytosine at PS4, guanine or cytosine at PS5, cytosine or thymine at PS6, adenine or guanine at PS7, cytosine or guanine at PS8, cytosine or thymine at PS9, cytosine or adenine at PS10, adenine or cytosine at PS11 and guanine or thymine at PS12. The invention also provides a collection of SCYA8 isogenes, referred to herein as a SCYA8 genome anthology.

In another embodiment, the invention provides a polynucleotide comprising a polymorphic variant of a reference sequence for a SCYA8 cDNA or a fragment thereof. The reference sequence comprises SEQ ID NO:2 (Fig.2) and the polymorphic cDNA comprises at least one polymorphism selected from the group consisting of thymine at a position corresponding to nucleotide 117, adenine at a position corresponding to nucleotide 201, cytosine at a position corresponding to nucleotide 235 and thymine at a position corresponding to nucleotide 250. A particularly preferred polymorphic cDNA variant comprises the coding sequence of a SCYA8 isogene defined by haplotypes 1-4, 6, 8, 9 and 11.

Polynucleotides complementary to these SCYA8 genomic and cDNA variants are also provided by the invention. It is believed that polymorphic variants of the SCYA8 gene will be useful in studying the expression and function of SCYA8, and in expressing SCYA8 protein for use in screening for candidate drugs to treat diseases related to SCYA8 activity.

In other embodiments, the invention provides a recombinant expression vector comprising one of the polymorphic genomic and cDNA variants operably linked to expression regulatory elements as well as a recombinant host cell transformed or transfected with the expression vector. The recombinant vector and host cell may be used to express SCYA8 for protein structure analysis and drug binding studies.

In yet another embodiment, the invention provides a polypeptide comprising a polymorphic variant of a reference amino acid sequence for the SCYA8 protein. The reference amino acid sequence comprises SEQ ID NO:3 (Fig.3) and the polymorphic variant comprises at least one variant amino acid selected from the group consisting of glutamine at a position corresponding to amino acid position 79 and phenylalanine at a position corresponding to amino acid position 84. A polymorphic variant of SCYA8 is useful in studying the effect of the variation on the biological activity of SCYA8 as well as on the binding affinity of candidate drugs targeting SCYA8 for the treatment of inflammatory diseases and HIV.

The present invention also provides antibodies that recognize and bind to the above polymorphic SCYA8 protein variant. Such antibodies can be utilized in a variety of diagnostic and prognostic formats and therapeutic methods.

The present invention also provides nonhuman transgenic animals comprising one or more of the SCYA8 polymorphic genomic variants described herein and methods for producing such animals. The transgenic animals are useful for studying expression of the SCYA8 isogenes *in vivo*, for *in vivo*

screening and testing of drugs targeted against SCYA8 protein, and for testing the efficacy of therapeutic agents and compounds for inflammatory diseases and HIV in a biological system.

The present invention also provides a computer system for storing and displaying polymorphism data determined for the SCYA8 gene. The computer system comprises a computer processing unit; a display; and a database containing the polymorphism data. The polymorphism data includes one or more of the following: the polymorphisms, the genotypes, the haplotypes, and the haplotype pairs identified for the SCYA8 gene in a reference population. In a preferred embodiment, the computer system is capable of producing a display showing SCYA8 haplotypes organized according to their evolutionary relationships.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 illustrates a reference sequence for the SCYA8 gene (Genaissance Reference No. 1097793; contiguous lines), with the start and stop positions of each region of coding sequence indicated with a bracket ([or]) and the numerical position below the sequence and the polymorphic site(s) and polymorphism(s) identified by Applicants in a reference population indicated by the variant nucleotide positioned below the polymorphic site in the sequence. SEQ ID NO:1 is equivalent to Figure 1, with the two alternative allelic variants of each polymorphic site indicated by the appropriate nucleotide symbol (R= G or A, Y= T or C, M= A or C, K= G or T, S= G or C, and W= A or T; WIPO standard ST.25). SEQ ID NO:64 is a modified version of SEQ ID NO:1 that shows the context sequence of each polymorphic site, PS1-PS12, in a uniform format to facilitate electronic searching. For each polymorphic site, SEQ ID NO:64 contains a block of 60 bases of the nucleotide sequence encompassing the centrally-located polymorphic site at the 30th position, followed by 60 bases of unspecified sequence to represent that each PS is separated by genomic sequence whose composition is defined elsewhere herein.

Figure 2 illustrates a reference sequence for the SCYA8 coding sequence (contiguous lines; SEQ ID NO:2), with the polymorphic site(s) and polymorphism(s) identified by Applicants in a reference population indicated by the variant nucleotide positioned below the polymorphic site in the sequence.

Figure 3 illustrates a reference sequence for the SCYA8 protein (contiguous lines; SEQ ID NO:3), with the variant amino acid(s) caused by the polymorphism(s) of Figure 2 positioned below the polymorphic site in the sequence.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

The present invention is based on the discovery of novel variants of the SCYA8 gene. As described in more detail below, the inventors herein discovered 11 isogenes of the SCYA8 gene by characterizing the SCYA8 gene found in genomic DNAs isolated from an Index Repository that contains immortalized cell lines from one chimpanzee and 93 human individuals. The human

individuals included a reference population of 79 unrelated individuals self-identified as belonging to one of four major population groups: Caucasian (21 individuals), African descent (20 individuals), Asian (20 individuals), or Hispanic/Latino (18 individuals). To the extent possible, the members of this reference population were organized into population subgroups by their self-identified ethnogeographic origin as shown in Table 1 below.

Table 1. Population Groups in the Index Repository

Population Group	Population Subgroup	No. of Individuals
African descent		20
	Sierra Leone	1
Asian		20
	Burma	1
	China	3
	Japan	6
	Korea	1
	Philippines	5
	Vietnam	4
Caucasian		21
	British Isles	3
	British Isles/Central	4
	British Isles/Eastern	1
	Central/Eastern	1
	Eastern	3
	Central/Mediterranean	1
	Mediterranean	2
	Scandinavian	2
Hispanic/Latino		18
	Caribbean	8
	Caribbean (Spanish Descent)	2
	Central American (Spanish Descent)	1
	Mexican American	4
	South American (Spanish Descent)	3

In addition, the Index Repository contains three unrelated indigenous American Indians (one from each of North, Central and South America), one three-generation Caucasian family (from the CEPH Utah cohort) and one two-generation African-American family.

The SCYA8 isogenes present in the human reference population are defined by haplotypes for 12 polymorphic sites in the SCYA8 gene, all of which are believed to be novel. The novel SCYA8 polymorphic sites identified by the inventors are referred to as PS1-PS12 to designate the order in which they are located in the gene (see Table 3 below). Using the genotypes identified in the Index Repository for PS1-PS12 and the methodology described in the Examples below, the inventors herein also determined the pair of haplotypes for the SCYA8 gene present in individual human members of this repository. The human genotypes and haplotypes found in the repository for the SCYA8 gene include those shown in Tables 4 and 5, respectively. The polymorphism and haplotype data disclosed herein are useful for validating whether SCYA8 is a suitable target for drugs to treat inflammatory

diseases and HIV, screening for such drugs and reducing bias in clinical trials of such drugs.

In the context of this disclosure, the following terms shall be defined as follows unless otherwise indicated:

Allele - A particular form of a genetic locus, distinguished from other forms by its particular nucleotide sequence.

Candidate Gene – A gene which is hypothesized to be responsible for a disease, condition, or the response to a treatment, or to be correlated with one of these.

Gene - A segment of DNA that contains all the information for the regulated biosynthesis of an RNA product, including promoters, exons, introns, and other untranslated regions that control expression.

Genotype – An unphased 5′ to 3′ sequence of nucleotide pair(s) found at one or more polymorphic sites in a locus on a pair of homologous chromosomes in an individual. As used herein, genotype includes a full-genotype and/or a sub-genotype as described below.

Full-genotype – The unphased 5′ to 3′ sequence of nucleotide pairs found at all polymorphic sites examined herein in a locus on a pair of homologous chromosomes in a single individual.

Sub-genotype – The unphased 5′ to 3′ sequence of nucleotides seen at a subset of the polymorphic sites examined herein in a locus on a pair of homologous chromosomes in a single individual.

Genotyping – A process for determining a genotype of an individual.

Haplotype – A 5′ to 3′ sequence of nucleotides found at one or more polymorphic sites in a locus on a single chromosome from a single individual. As used herein, haplotype includes a full-haplotype and/or a sub-haplotype as described below.

Full-haplotype – The 5′ to 3′ sequence of nucleotides found at all polymorphic sites examined herein in a locus on a single chromosome from a single individual.

Sub-haplotype – The 5′ to 3′ sequence of nucleotides seen at a subset of the polymorphic sites examined herein in a locus on a single chromosome from a single individual.

Haplotype pair – The two haplotypes found for a locus in a single individual.

Haplotyping – A process for determining one or more haplotypes in an individual and includes use of family pedigrees, molecular techniques and/or statistical inference.

Haplotype data - Information concerning one or more of the following for a specific gene: a listing of the haplotype pairs in each individual in a population; a listing of the different haplotypes in a population; frequency of each haplotype in that or other populations, and any known associations between one or more haplotypes and a trait.

Isoform – A particular form of a gene, mRNA, cDNA or the protein encoded thereby, distinguished from other forms by its particular sequence and/or structure.

Isogene – One of the isoforms (e.g., alleles) of a gene found in a population. An isogene (or allele) contains all of the polymorphisms present in the particular isoform of the gene.

Isolated – As applied to a biological molecule such as RNA, DNA, oligonucleotide, or protein, isolated means the molecule is substantially free of other biological molecules such as nucleic acids, proteins, lipids, carbohydrates, or other material such as cellular debris and growth media. Generally, the term "isolated" is not intended to refer to a complete absence of such material or to absence of water, buffers, or salts, unless they are present in amounts that substantially interfere with the methods of the present invention.

Locus - A location on a chromosome or DNA molecule corresponding to a gene or a physical or phenotypic feature, where physical features include polymorphic sites.

Naturally-occurring – A term used to designate that the object it is applied to, e.g., naturally-occurring polynucleotide or polypeptide, can be isolated from a source in nature and which has not been intentionally modified by man.

Nucleotide pair – The nucleotides found at a polymorphic site on the two copies of a chromosome from an individual.

Phased – As applied to a sequence of nucleotide pairs for two or more polymorphic sites in a locus, phased means the combination of nucleotides present at those polymorphic sites on a single copy of the locus is known.

Polymorphic site (PS) – A position on a chromosome or DNA molecule at which at least two alternative sequences are found in a population.

Polymorphic variant – A gene, mRNA, cDNA, polypeptide or peptide whose nucleotide or amino acid sequence varies from a reference sequence due to the presence of a polymorphism in the gene.

Polymorphism – The sequence variation observed in an individual at a polymorphic site. Polymorphisms include nucleotide substitutions, insertions, deletions and microsatellites and may, but need not, result in detectable differences in gene expression or protein function.

Polymorphism data – Information concerning one or more of the following for a specific gene: location of polymorphic sites; sequence variation at those sites; frequency of polymorphisms in one or more populations; the different genotypes and/or haplotypes determined for the gene; frequency of one or more of these genotypes and/or haplotypes in one or more populations; any known association(s) between a trait and a genotype or a haplotype for the gene.

Polymorphism Database – A collection of polymorphism data arranged in a systematic or methodical way and capable of being individually accessed by electronic or other means.

Polynucleotide – A nucleic acid molecule comprised of single-stranded RNA or DNA or comprised of complementary, double-stranded DNA.

Population Group – A group of individuals sharing a common ethnogeographic origin.

Reference Population – A group of subjects or individuals who are predicted to be representative of the genetic variation found in the general population. Typically, the reference population represents the genetic variation in the population at a certainty level of at least 85%,

preferably at least 90%, more preferably at least 95% and even more preferably at least 99%.

Single Nucleotide Polymorphism (SNP) – Typically, the specific pair of nucleotides observed at a single polymorphic site. In rare cases, three or four nucleotides may be found.

Subject – A human individual whose genotypes or haplotypes or response to treatment or disease state are to be determined.

Treatment - A stimulus administered internally or externally to a subject.

Unphased – As applied to a sequence of nucleotide pairs for two or more polymorphic sites in a locus, unphased means the combination of nucleotides present at those polymorphic sites on a single copy of the locus is not known.

As discussed above, information on the identity of genotypes and haplotypes for the SCYA8 gene of any particular individual as well as the frequency of such genotypes and haplotypes in any particular population of individuals is useful for a variety of drug discovery and development applications. Thus, the invention also provides compositions and methods for detecting the novel SCYA8 polymorphisms, haplotypes and haplotype pairs identified herein.

The compositions comprise at least one oligonucleotide for detecting the variant nucleotide or nucleotide pair located at a novel SCYA8 polymorphic site in one copy or two copies of the SCYA8 gene. Such oligonucleotides are referred to herein as SCYA8 haplotyping oligonucleotides or genotyping oligonucleotides, respectively, and collectively as SCYA8 oligonucleotides. In one embodiment, a SCYA8 haplotyping or genotyping oligonucleotide is a probe or primer capable of hybridizing to a target region that contains, or that is located close to, one of the novel polymorphic sites described herein.

As used herein, the term "oligonucleotide" refers to a polynucleotide molecule having less than about 100 nucleotides. A preferred oligonucleotide of the invention is 10 to 35 nucleotides long. More preferably, the oligonucleotide is between 15 and 30, and most preferably, between 20 and 25 nucleotides in length. The exact length of the oligonucleotide will depend on many factors that are routinely considered and practiced by the skilled artisan. The oligonucleotide may be comprised of any phosphorylation state of ribonucleotides, deoxyribonucleotides, and acyclic nucleotide derivatives, and other functionally equivalent derivatives. Alternatively, oligonucleotides may have a phosphate-free backbone, which may be comprised of linkages such as carboxymethyl, acetamidate, carbamate, polyamide (peptide nucleic acid (PNA)) and the like (Varma, R. in Molecular Biology and Biotechnology, A Comprehensive Desk Reference, Ed. R. Meyers, VCH Publishers, Inc. (1995), pages 617-620). Oligonucleotides of the invention may be prepared by chemical synthesis using any suitable methodology known in the art, or may be derived from a biological sample, for example, by restriction digestion. The oligonucleotides may be labeled, according to any technique known in the art, including use of radiolabels, fluorescent labels, enzymatic labels, proteins, haptens, antibodies, sequence tags and the like.

Haplotyping or genotyping oligonucleotides of the invention must be capable of specifically

hybridizing to a target region of a SCYA8 polynucleotide. Preferably, the target region is located in a SCYA8 isogene. As used herein, specific hybridization means the oligonucleotide forms an anti-parallel double-stranded structure with the target region under certain hybridizing conditions, while failing to form such a structure when incubated with another region in the SCYA8 polynucleotide or with a non-SCYA8 polynucleotide under the same hybridizing conditions. Preferably, the oligonucleotide specifically hybridizes to the target region under conventional high stringency conditions. The skilled artisan can readily design and test oligonucleotide probes and primers suitable for detecting polymorphisms in the SCYA8 gene using the polymorphism information provided herein in conjunction with the known sequence information for the SCYA8 gene and routine techniques.

A nucleic acid molecule such as an oligonucleotide or polynucleotide is said to be a "perfect" or "complete" complement of another nucleic acid molecule if every nucleotide of one of the molecules is complementary to the nucleotide at the corresponding position of the other molecule. A nucleic acid molecule is "substantially complementary" to another molecule if it hybridizes to that molecule with sufficient stability to remain in a duplex form under conventional low-stringency conditions. Conventional hybridization conditions are described, for example, by Sambrook J. et al., in Molecular Cloning, A Laboratory Manual, 2nd Edition, Cold Spring Harbor Press, Cold Spring Harbor, NY (1989) and by Haymes, B.D. et al. in Nucleic Acid Hybridization, A Practical Approach, IRL Press, Washington, D.C. (1985). While perfectly complementary oligonucleotides are preferred for detecting polymorphisms, departures from complete complementarity are contemplated where such departures do not prevent the molecule from specifically hybridizing to the target region. For example, an oligonucleotide primer may have a non-complementary fragment at its 5′ end, with the remainder of the primer being complementary to the target region. Alternatively, non-complementary nucleotides may be interspersed into the probe or primer as long as the resulting probe or primer is still capable of specifically hybridizing to the target region.

Preferred haplotyping or genotyping oligonucleotides of the invention are allele-specific oligonucleotides. As used herein, the term allele-specific oligonucleotide (ASO) means an oligonucleotide that is able, under sufficiently stringent conditions, to hybridize specifically to one allele of a gene, or other locus, at a target region containing a polymorphic site while not hybridizing to the corresponding region in another allele(s). As understood by the skilled artisan, allele-specificity will depend upon a variety of readily optimized stringency conditions, including salt and formamide concentrations, as well as temperatures for both the hybridization and washing steps. Examples of hybridization and washing conditions typically used for ASO probes are found in Kogan et al., "Genetic Prediction of Hemophilia A" in PCR Protocols, A Guide to Methods and Applications, Academic Press, 1990 and Ruaño et al., 87 *Proc. Natl. Acad. Sci. USA* 6296-6300, 1990. Typically, an ASO will be perfectly complementary to one allele while containing a single mismatch for another allele.

Allele-specific oligonucleotides of the invention include ASO probes and ASO primers. ASO

probes which usually provide good discrimination between different alleles are those in which a central position of the oligonucleotide probe aligns with the polymorphic site in the target region (e.g., approximately the 7th or 8th position in a 15mer, the 8th or 9th position in a 16mer, and the 10th or 11th position in a 20mer). An ASO primer of the invention has a 3′ terminal nucleotide, or preferably a 3′ penultimate nucleotide, that is complementary to only one nucleotide of a particular SNP, thereby acting as a primer for polymerase-mediated extension only if the allele containing that nucleotide is present. ASO probes and primers hybridizing to either the coding or noncoding strand are contemplated by the invention. ASO probes and primers listed below use the appropriate nucleotide symbol (R= G or A, Y= T or C, M= A or C, K= G or T, S= G or C, and W= A or T; WIPO standard ST.25) at the position of the polymorphic site to represent that the ASO contains either of the two alternative allelic variants observed at that polymorphic site.

A preferred ASO probe for detecting SCYA8 gene polymorphisms comprises a nucleotide sequence, listed 5′ to 3′, selected from the group consisting of:

```
CCCTCTGYAACAGTA   (SEQ ID NO:4) and its complement,
AGGGGTARGCCCTTG   (SEQ ID NO:5) and its complement,
CTGGTGCYTACTCAG   (SEQ ID NO:6) and its complement,
TGTGACCMTGCCCAG   (SEQ ID NO:7) and its complement,
TGCCCAGSCTCTCTG   (SEQ ID NO:8) and its complement,
GAGCCACYGAGGAGC   (SEQ ID NO:9) and its complement,
TGAGAACRACCCAGA   (SEQ ID NO:10) and its complement,
GGTGGGTSCTAAATG   (SEQ ID NO:11) and its complement,
CAGTTTCYATTCCAA   (SEQ ID NO:12) and its complement,
CCAACATMCAATGTC   (SEQ ID NO:13) and its complement,
CAAGACCMAACGGGG   (SEQ ID NO:14) and its complement,and
CAAGGAGKTCTGTGC   (SEQ ID NO:15) and its complement.
```

A preferred ASO primer for detecting SCYA8 gene polymorphisms comprises a nucleotide sequence, listed 5′ to 3′, selected from the group consisting of:

```
TTCTTACCCTCTGYA (SEQ ID NO:16); GAATCTTACTGTTRC (SEQ ID NO:17);
CCACAGAGGGGTARG (SEQ ID NO:18); GAGAGCCAAGGGCYT (SEQ ID NO:19);
CAGGATCTGGTGCYT (SEQ ID NO:20); AATATGCTGAGTARG (SEQ ID NO:21);
CTTGATTGTGACCMT (SEQ ID NO:22); GAGAGCCTGGGCAKG (SEQ ID NO:23);
TGACCATGCCCAGSC (SEQ ID NO:24); AGGGAGCAGAGAGSC (SEQ ID NO:25);
CAGGCAGAGCCACYG (SEQ ID NO:26); CTCTCTGCTCCTCRG (SEQ ID NO:27);
AGAGGTTGAGAACRA (SEQ ID NO:28); AAGGTTTCTGGGTYG (SEQ ID NO:29);
ACTTACGGTGGGTSC (SEQ ID NO:30); ATGAGACATTTAGSA (SEQ ID NO:31);
CAGATTCAGTTTCYA (SEQ ID NO:32); AGGTGATTGGAATRG (SEQ ID NO:33);
GAATCACCAACATMC (SEQ ID NO:34); CCTTGGGACATTGKA (SEQ ID NO:35);
CAGCTTCAAGACCMA (SEQ ID NO:36); TCCTTGCCCCGTTKG (SEQ ID NO:37);
ACGGGGCAAGGAGKT (SEQ ID NO:38); and GGGTCAGCACAGAMC (SEQ ID NO:39).
```

Other oligonucleotides of the invention hybridize to a target region located one to several nucleotides downstream of one of the novel polymorphic sites identified herein. Such oligonucleotides are useful in polymerase-mediated primer extension methods for detecting one of the novel polymorphisms described herein and therefore such oligonucleotides are referred to herein as "primer-

extension oligonucleotides". In a preferred embodiment, the 3′-terminus of a primer-extension oligonucleotide is a deoxynucleotide complementary to the nucleotide located immediately adjacent to the polymorphic site.

A particularly preferred oligonucleotide primer for detecting SCYA8 gene polymorphisms by primer extension terminates in a nucleotide sequence, listed 5′ to 3′, selected from the group consisting of:

```
TTACCCTCTG   (SEQ ID NO:40); TCTTACTGTT   (SEQ ID NO:41);
CAGAGGGGTA   (SEQ ID NO:42); AGCCAAGGGC   (SEQ ID NO:43);
GATCTGGTGC   (SEQ ID NO:44); ATGCTGAGTA   (SEQ ID NO:45);
GATTGTGACC   (SEQ ID NO:46); AGCCTGGGCA   (SEQ ID NO:47);
CCATGCCCAG   (SEQ ID NO:48); GAGCAGAGAG   (SEQ ID NO:49);
GCAGAGCCAC   (SEQ ID NO:50); TCTGCTCCTC   (SEQ ID NO:51);
GGTTGAGAAC   (SEQ ID NO:52); GTTTCTGGGT   (SEQ ID NO:53);
TACGGTGGGT   (SEQ ID NO:54); AGACATTTAG   (SEQ ID NO:55);
ATTCAGTTTC   (SEQ ID NO:56); TGATTGGAAT   (SEQ ID NO:57);
TCACCAACAT   (SEQ ID NO:58); TGGGACATTG   (SEQ ID NO:59);
CTTCAAGACC   (SEQ ID NO:60); TTGCCCCGTT   (SEQ ID NO:61);
GGGCAAGGAG   (SEQ ID NO:62); and TCAGCACAGA   (SEQ ID NO:63).
```

In some embodiments, a composition contains two or more differently labeled SCYA8 oligonucleotides for simultaneously probing the identity of nucleotides or nucleotide pairs at two or more polymorphic sites. It is also contemplated that primer compositions may contain two or more sets of allele-specific primer pairs to allow simultaneous targeting and amplification of two or more regions containing a polymorphic site.

SCYA8 oligonucleotides of the invention may also be immobilized on or synthesized on a solid surface such as a microchip, bead, or glass slide (see, e.g., WO 98/20020 and WO 98/20019). Such immobilized oligonucleotides may be used in a variety of polymorphism detection assays, including but not limited to probe hybridization and polymerase extension assays. Immobilized SCYA8 oligonucleotides of the invention may comprise an ordered array of oligonucleotides designed to rapidly screen a DNA sample for polymorphisms in multiple genes at the same time.

In another embodiment, the invention provides a kit comprising at least two SCYA8 oligonucleotides packaged in separate containers. The kit may also contain other components such as hybridization buffer (where the oligonucleotides are to be used as a probe) packaged in a separate container. Alternatively, where the oligonucleotides are to be used to amplify a target region, the kit may contain, packaged in separate containers, a polymerase and a reaction buffer optimized for primer extension mediated by the polymerase, such as PCR.

The above described oligonucleotide compositions and kits are useful in methods for genotyping and/or haplotyping the SCYA8 gene in an individual. As used herein, the terms "SCYA8 genotype" and "SCYA8 haplotype" mean the genotype or haplotype contains the nucleotide pair or nucleotide, respectively, that is present at one or more of the novel polymorphic sites described herein and may optionally also include the nucleotide pair or nucleotide present at one or more additional

polymorphic sites in the SCYA8 gene. The additional polymorphic sites may be currently known polymorphic sites or sites that are subsequently discovered.

One embodiment of a genotyping method of the invention involves isolating from the individual a nucleic acid sample comprising the two copies of the SCYA8 gene, mRNA transcripts thereof or cDNA copies thereof, or a fragment of any of the foregoing, that are present in the individual, and determining the identity of the nucleotide pair at one or more polymorphic sites selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11 and PS12 in the two copies to assign a SCYA8 genotype to the individual. As will be readily understood by the skilled artisan, the two "copies" of a gene, mRNA or cDNA (or fragment of such SCYA8 molecules) in an individual may be the same allele or may be different alleles. In another embodiment, a genotyping method of the invention comprises determining the identity of the nucleotide pair at each of PS1-PS12.

Typically, the nucleic acid sample is isolated from a biological sample taken from the individual, such as a blood sample or tissue sample. Suitable tissue samples include whole blood, semen, saliva, tears, urine, fecal material, sweat, buccal, skin and hair. The nucleic acid sample may be comprised of genomic DNA, mRNA, or cDNA and, in the latter two cases, the biological sample must be obtained from a tissue in which the SCYA8 gene is expressed. Furthermore it will be understood by the skilled artisan that mRNA or cDNA preparations would not be used to detect polymorphisms located in introns or in 5′ and 3′ untranslated regions if not present in the mRNA or cDNA. If a SCYA8 gene fragment is isolated, it must contain the polymorphic site(s) to be genotyped.

One embodiment of a haplotyping method of the invention comprises isolating from the individual a nucleic acid sample containing only one of the two copies of the SCYA8 gene, mRNA or cDNA, or a fragment of such SCYA8 molecules, that is present in the individual and determining in that copy the identity of the nucleotide at one or more polymorphic sites selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11 and PS12 in that copy to assign a SCYA8 haplotype to the individual.

The nucleic acid used in the above haplotyping methods of the invention may be isolated using any method capable of separating the two copies of the SCYA8 gene or fragment such as one of the methods described above for preparing SCYA8 isogenes, with targeted *in vivo* cloning being the preferred approach. As will be readily appreciated by those skilled in the art, any individual clone will typically only provide haplotype information on one of the two SCYA8 gene copies present in an individual. If haplotype information is desired for the individual's other copy, additional SCYA8 clones will usually need to be examined. Typically, at least five clones should be examined to have more than a 90% probability of haplotyping both copies of the SCYA8 gene in an individual. In some cases, however, once the haplotype for one SCYA8 allele is directly determined, the haplotype for the other allele may be inferred if the individual has a known genotype for the polymorphic sites of interest or if the haplotype frequency or haplotype pair frequency for the individual's population group

is known. In a particularly preferred embodiment, the nucleotide at each of PS1-PS12 is identified.

In another embodiment, the haplotyping method comprises determining whether an individual has one or more of the SCYA8 haplotypes shown in Table 5. This can be accomplished by identifying, for one or both copies of the individual's SCYA8 gene, the phased sequence of nucleotides present at each of PS1-PS12. This identifying step does not necessarily require that each of PS1-PS12 be directly examined. Typically only a subset of PS1-PS12 will need to be directly examined to assign to an individual one or more of the haplotypes shown in Table 5. This is because at least one polymorphic site in a gene is frequently in strong linkage disequilibrium with one or more other polymorphic sites in that gene (Drysdale, CM et al. 2000 *PNAS* **97**:10483-10488; Rieder MJ et al. 1999 *Nature Genetics* **22**:59-62). Two sites are said to be in linkage disequilibrium if the presence of a particular variant at one site enhances the predictability of another variant at the second site (Stephens, JC 1999, *Mol. Diag.* **4**:309-317). Techniques for determining whether any two polymorphic sites are in linkage disequilibrium are well-known in the art (Weir B.S. 1996 *Genetic Data Analysis II*, Sinauer Associates, Inc. Publishers, Sunderland, MA).

In another embodiment of a haplotyping method of the invention, a SCYA8 haplotype pair is determined for an individual by identifying the phased sequence of nucleotides at one or more polymorphic sites selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11 and PS12 in each copy of the SCYA8 gene that is present in the individual. In a particularly preferred embodiment, the haplotyping method comprises identifying the phased sequence of nucleotides at each of PS1-PS12 in each copy of the SCYA8 gene.

When haplotyping both copies of the gene, the identifying step is preferably performed with each copy of the gene being placed in separate containers. However, it is also envisioned that if the two copies are labeled with different tags, or are otherwise separately distinguishable or identifiable, it could be possible in some cases to perform the method in the same container. For example, if first and second copies of the gene are labeled with different first and second fluorescent dyes, respectively, and an allele-specific oligonucleotide labeled with yet a third different fluorescent dye is used to assay the polymorphic site(s), then detecting a combination of the first and third dyes would identify the polymorphism in the first gene copy while detecting a combination of the second and third dyes would identify the polymorphism in the second gene copy.

In both the genotyping and haplotyping methods, the identity of a nucleotide (or nucleotide pair) at a polymorphic site(s) may be determined by amplifying a target region(s) containing the polymorphic site(s) directly from one or both copies of the SCYA8 gene, or a fragment thereof, and the sequence of the amplified region(s) determined by conventional methods. It will be readily appreciated by the skilled artisan that only one nucleotide will be detected at a polymorphic site in individuals who are homozygous at that site, while two different nucleotides will be detected if the individual is heterozygous for that site. The polymorphism may be identified directly, known as positive-type identification, or by inference, referred to as negative-type identification. For example,

where a SNP is known to be guanine and cytosine in a reference population, a site may be positively determined to be either guanine or cytosine for an individual homozygous at that site, or both guanine and cytosine, if the individual is heterozygous at that site. Alternatively, the site may be negatively determined to be not guanine (and thus cytosine/cytosine) or not cytosine (and thus guanine/guanine).

The target region(s) may be amplified using any oligonucleotide-directed amplification method, including but not limited to polymerase chain reaction (PCR) (U.S. Patent No. 4,965,188), ligase chain reaction (LCR) (Barany et al., *Proc. Natl. Acad. Sci. USA* 88:189-193, 1991; WO90/01069), and oligonucleotide ligation assay (OLA) (Landegren et al., *Science* 241:1077-1080, 1988). Other known nucleic acid amplification procedures may be used to amplify the target region including transcription-based amplification systems (U.S. Patent No. 5,130,238; EP 329,822; U.S. Patent No. 5,169,766, WO89/06700) and isothermal methods (Walker et al., *Proc. Natl. Acad. Sci. USA* **89**:392-396, 1992).

A polymorphism in the target region may also be assayed before or after amplification using one of several hybridization-based methods known in the art. Typically, allele-specific oligonucleotides are utilized in performing such methods. The allele-specific oligonucleotides may be used as differently labeled probe pairs, with one member of the pair showing a perfect match to one variant of a target sequence and the other member showing a perfect match to a different variant. In some embodiments, more than one polymorphic site may be detected at once using a set of allele-specific oligonucleotides or oligonucleotide pairs. Preferably, the members of the set have melting temperatures within 5°C, and more preferably within 2°C, of each other when hybridizing to each of the polymorphic sites being detected.

Hybridization of an allele-specific oligonucleotide to a target polynucleotide may be performed with both entities in solution, or such hybridization may be performed when either the oligonucleotide or the target polynucleotide is covalently or noncovalently affixed to a solid support. Attachment may be mediated, for example, by antibody-antigen interactions, poly-L-Lys, streptavidin or avidin-biotin, salt bridges, hydrophobic interactions, chemical linkages, UV cross-linking baking, etc. Allele-specific oligonucleotides may be synthesized directly on the solid support or attached to the solid support subsequent to synthesis. Solid-supports suitable for use in detection methods of the invention include substrates made of silicon, glass, plastic, paper and the like, which may be formed, for example, into wells (as in 96-well plates), slides, sheets, membranes, fibers, chips, dishes, and beads. The solid support may be treated, coated or derivatized to facilitate the immobilization of the allele-specific oligonucleotide or target nucleic acid.

The genotype or haplotype for the SCYA8 gene of an individual may also be determined by hybridization of a nucleic acid sample containing one or both copies of the gene, mRNA, cDNA or fragment(s) thereof, to nucleic acid arrays and subarrays such as described in WO 95/11995. The arrays would contain a battery of allele-specific oligonucleotides representing each of the polymorphic sites to be included in the genotype or haplotype.

The identity of polymorphisms may also be determined using a mismatch detection technique, including but not limited to the RNase protection method using riboprobes (Winter et al., *Proc. Natl. Acad. Sci. USA* 82:7575, 1985; Meyers et al., *Science* 230:1242, 1985) and proteins which recognize nucleotide mismatches, such as the E. coli mutS protein (Modrich, P. *Ann. Rev. Genet.* 25:229-253, 1991). Alternatively, variant alleles can be identified by single strand conformation polymorphism (SSCP) analysis (Orita et al., *Genomics* 5:874-879, 1989; Humphries et al., in Molecular Diagnosis of Genetic Diseases, R. Elles, ed., pp. 321-340, 1996) or denaturing gradient gel electrophoresis (DGGE) (Wartell et al., *Nucl. Acids Res.* 18:2699-2706, 1990; Sheffield et al., *Proc. Natl. Acad. Sci. USA* 86:232-236, 1989).

A polymerase-mediated primer extension method may also be used to identify the polymorphism(s). Several such methods have been described in the patent and scientific literature and include the "Genetic Bit Analysis" method (WO92/15712) and the ligase/polymerase mediated genetic bit analysis (U.S. Patent 5,679,524. Related methods are disclosed in WO91/02087, WO90/09455, WO95/17676, U.S. Patent Nos. 5,302,509, and 5,945,283. Extended primers containing a polymorphism may be detected by mass spectrometry as described in U.S. Patent No. 5,605,798. Another primer extension method is allele-specific PCR (Ruaño et al., *Nucl. Acids Res.* 17:8392, 1989; Ruaño et al., *Nucl. Acids Res.* 19, 6877-6882, 1991; WO 93/22456; Turki et al., *J. Clin. Invest.* 95:1635-1641, 1995). In addition, multiple polymorphic sites may be investigated by simultaneously amplifying multiple regions of the nucleic acid using sets of allele-specific primers as described in Wallace et al. (WO89/10414).

In addition, the identity of the allele(s) present at any of the novel polymorphic sites described herein may be indirectly determined by haplotyping or genotyping another polymorphic site that is in linkage disequilibrium with the polymorphic site that is of interest. Polymorphic sites in linkage disequilibrium with the presently disclosed polymorphic sites may be located in regions of the gene or in other genomic regions not examined herein. Detection of the allele(s) present at a polymorphic site in linkage disequilibrium with the novel polymorphic sites described herein may be performed by, but is not limited to, any of the above-mentioned methods for detecting the identity of the allele at a polymorphic site.

In another aspect of the invention, an individual's SCYA8 haplotype pair is predicted from its SCYA8 genotype using information on haplotype pairs known to exist in a reference population. In its broadest embodiment, the haplotyping prediction method comprises identifying a SCYA8 genotype for the individual at two or more SCYA8 polymorphic sites described herein, accessing data containing SCYA8 haplotype pairs identified in a reference population, and assigning a haplotype pair to the individual that is consistent with the genotype data. In one embodiment, the reference haplotype pairs include the SCYA8 haplotype pairs shown in Table 4. The SCYA8 haplotype pair can be assigned by comparing the individual's genotype with the genotypes corresponding to the haplotype pairs known to exist in the general population or in a specific population group, and determining which haplotype pair

is consistent with the genotype of the individual. In some embodiments, comparison of the genotype of the individual to the haplotype pairs identified in a reference population and determination of which haplotype pair is consistent with the genotype of the individual may be performed by visual inspection (for example, by consulting Table 4). When the genotype of the individual is consistent with more than one haplotype pair, haplotype pair frequency data (such as that presented in Table 7) may be used to determine which of these haplotype pairs is most likely to be present in the individual. This determination may also be performed in some embodiments by visual inspection upon consulting Table 7. If a particular SCYA8 haplotype pair consistent with the genotype of the individual is more frequent in the reference population than others consistent with the genotype, then that haplotype pair with the highest frequency is the most likely to be present in the individual. In other embodiments, the comparison may be made by a computer-implemented algorithm with the genotype of the individual and the reference haplotype data stored in computer-readable formats. For example, as described in PCT/US01/12831, filed April 18, 2001, one computer-implemented algorithm to perform this comparison entails enumerating all possible haplotype pairs which are consistent with the genotype, accessing data containing SCYA8 haplotype pairs frequency data determined in a reference population to determine a probability that the individual has a possible haplotype pair, and analyzing the determined probabilities to assign a haplotype pair to the individual.

Generally, the reference population should be composed of randomly-selected individuals representing the major ethnogeographic groups of the world. A preferred reference population for use in the methods of the present invention comprises an approximately equal number of individuals from Caucasian, African-descent, Asian and Hispanic-Latino population groups with the minimum number of each group being chosen based on how rare a haplotype one wants to be guaranteed to see. For example, if one wants to have a q% chance of not missing a haplotype that exists in the population at a p% frequency of occurring in the reference population, the number of individuals (n) who must be sampled is given by 2n=log(1-q)/log(1-p) where p and q are expressed as fractions. A preferred reference population allows the detection of any haplotype whose frequency is at least 10% with about 99% certainty and comprises about 20 unrelated individuals from each of the four population groups named above. A particularly preferred reference population includes a 3-generation family representing one or more of the four population groups to serve as controls for checking quality of haplotyping procedures.

In a preferred embodiment, the haplotype frequency data for each ethnogeographic group is examined to determine whether it is consistent with Hardy-Weinberg equilibrium. Hardy-Weinberg equilibrium (D.L. Hartl et al., Principles of Population Genomics, Sinauer Associates (Sunderland, MA), 3rd Ed., 1997) postulates that the frequency of finding the haplotype pair H_1/H_2 is equal to

$$p_{H-W}(H_1/H_2) = 2p(H_1)p(H_2) \text{ if } H_1 \neq H_2 \text{ and } p_{H-W}(H_1/H_2) = p(H_1)p(H_2) \text{ if } H_1 = H_2.$$

A statistically significant difference between the observed and expected haplotype frequencies could be due to one or more factors including significant inbreeding in the population group, strong selective

pressure on the gene, sampling bias, and/or errors in the genotyping process. If large deviations from Hardy-Weinberg equilibrium are observed in an ethnogeographic group, the number of individuals in that group can be increased to see if the deviation is due to a sampling bias. If a larger sample size does not reduce the difference between observed and expected haplotype pair frequencies, then one may wish to consider haplotyping the individual using a direct haplotyping method such as, for example, CLASPER System™ technology (U.S. Patent No. 5,866,404), single molecule dilution, or allele-specific long-range PCR (Michalotos-Beloin et al., *Nucleic Acids Res.* 24:4841-4843, 1996).

In one embodiment of this method for predicting a SCYA8 haplotype pair for an individual, the assigning step involves performing the following analysis. First, each of the possible haplotype pairs is compared to the haplotype pairs in the reference population. Generally, only one of the haplotype pairs in the reference population matches a possible haplotype pair and that pair is assigned to the individual. Occasionally, only one haplotype represented in the reference haplotype pairs is consistent with a possible haplotype pair for an individual, and in such cases the individual is assigned a haplotype pair containing this known haplotype and a new haplotype derived by subtracting the known haplotype from the possible haplotype pair. Alternatively, the haplotype pair in an individual may be predicted from the individual's genotype for that gene using reported methods (e.g., Clark et al. 1990 *Mol Bio Evol* 7:111-22; copending PCT/US01/12831 filed April 18, 2001) or through a commercial haplotyping service such as offered by Genaissance Pharmaceuticals, Inc. (New Haven, CT). In rare cases, either no haplotypes in the reference population are consistent with the possible haplotype pairs, or alternatively, multiple reference haplotype pairs are consistent with the possible haplotype pairs. In such cases, the individual is preferably haplotyped using a direct molecular haplotyping method such as, for example, CLASPER System™ technology (U.S. Patent No. 5,866,404), SMD, or allele-specific long-range PCR (Michalotos-Beloin et al., *supra*).

The invention also provides a method for determining the frequency of a SCYA8 genotype, haplotype, or haplotype pair in a population. The method comprises, for each member of the population, determining the genotype or the haplotype pair for the novel SCYA8 polymorphic sites described herein, and calculating the frequency any particular genotype, haplotype, or haplotype pair is found in the population. The population may be e.g., a reference population, a family population, a same gender population, a population group, or a trait population (e.g., a group of individuals exhibiting a trait of interest such as a medical condition or response to a therapeutic treatment).

In another aspect of the invention, frequency data for SCYA8 genotypes, haplotypes, and/or haplotype pairs are determined in a reference population and used in a method for identifying an association between a trait and a SCYA8 genotype, haplotype, or haplotype pair. The trait may be any detectable phenotype, including but not limited to susceptibility to a disease or response to a treatment. In one embodiment, the method involves obtaining data on the frequency of the genotype(s), haplotype(s), or haplotype pair(s) of interest in a reference population as well as in a population exhibiting the trait. Frequency data for one or both of the reference and trait populations may be

obtained by genotyping or haplotyping each individual in the populations using one or more of the methods described above. The haplotypes for the trait population may be determined directly or, alternatively, by a predictive genotype to haplotype approach as described above. In another embodiment, the frequency data for the reference and/or trait populations is obtained by accessing previously determined frequency data, which may be in written or electronic form. For example, the frequency data may be present in a database that is accessible by a computer. Once the frequency data is obtained, the frequencies of the genotype(s), haplotype(s), or haplotype pair(s) of interest in the reference and trait populations are compared. In a preferred embodiment, the frequencies of all genotypes, haplotypes, and/or haplotype pairs observed in the populations are compared. If a particular SCYA8 genotype, haplotype, or haplotype pair is more frequent in the trait population than in the reference population at a statistically significant amount, then the trait is predicted to be associated with that SCYA8 genotype, haplotype or haplotype pair. Preferably, the SCYA8 genotype, haplotype, or haplotype pair being compared in the trait and reference populations is selected from the full-genotypes and full-haplotypes shown in Tables 4 and 5, or from sub-genotypes and sub-haplotypes derived from these genotypes and haplotypes.

In a preferred embodiment of the method, the trait of interest is a clinical response exhibited by a patient to some therapeutic treatment, for example, response to a drug targeting SCYA8 or response to a therapeutic treatment for a medical condition. As used herein, "medical condition" includes but is not limited to any condition or disease manifested as one or more physical and/or psychological symptoms for which treatment is desirable, and includes previously and newly identified diseases and other disorders. As used herein the term "clinical response" means any or all of the following: a quantitative measure of the response, no response, and/or adverse response (i.e., side effects).

In order to deduce a correlation between clinical response to a treatment and a SCYA8 genotype, haplotype, or haplotype pair, it is necessary to obtain data on the clinical responses exhibited by a population of individuals who received the treatment, hereinafter the "clinical population". This clinical data may be obtained by analyzing the results of a clinical trial that has already been run and/or the clinical data may be obtained by designing and carrying out one or more new clinical trials. As used herein, the term "clinical trial" means any research study designed to collect clinical data on responses to a particular treatment, and includes but is not limited to phase I, phase II and phase III clinical trials. Standard methods are used to define the patient population and to enroll subjects.

It is preferred that the individuals included in the clinical population have been graded for the existence of the medical condition of interest. This is important in cases where the symptom(s) being presented by the patients can be caused by more than one underlying condition, and where treatment of the underlying conditions are not the same. An example of this would be where patients experience breathing difficulties that are due to either asthma or respiratory infections. If both sets were treated with an asthma medication, there would be a spurious group of apparent non-responders that did not

actually have asthma. These people would affect the ability to detect any correlation between haplotype and treatment outcome. This grading of potential patients could employ a standard physical exam or one or more lab tests. Alternatively, grading of patients could use haplotyping for situations where there is a strong correlation between haplotype pair and disease susceptibility or severity.

The therapeutic treatment of interest is administered to each individual in the trial population and each individual's response to the treatment is measured using one or more predetermined criteria. It is contemplated that in many cases, the trial population will exhibit a range of responses and that the investigator will choose the number of responder groups (e.g., low, medium, high) made up by the various responses. In addition, the SCYA8 gene for each individual in the trial population is genotyped and/or haplotyped, which may be done before or after administering the treatment.

After both the clinical and polymorphism data have been obtained, correlations between individual response and SCYA8 genotype or haplotype content are created. Correlations may be produced in several ways. In one method, individuals are grouped by their SCYA8 genotype or haplotype (or haplotype pair) (also referred to as a polymorphism group), and then the averages and standard deviations of clinical responses exhibited by the members of each polymorphism group are calculated.

These results are then analyzed to determine if any observed variation in clinical response between polymorphism groups is statistically significant. Statistical analysis methods which may be used are described in L.D. Fisher and G. vanBelle, "Biostatistics: A Methodology for the Health Sciences", Wiley-Interscience (New York) 1993. This analysis may also include a regression calculation of which polymorphic sites in the SCYA8 gene give the most significant contribution to the differences in phenotype. One regression model useful in the invention is described in WO 01/01218, entitled "Methods for Obtaining and Using Haplotype Data".

A second method for finding correlations between SCYA8 haplotype content and clinical responses uses predictive models based on error-minimizing optimization algorithms. One of many possible optimization algorithms is a genetic algorithm (R. Judson, "Genetic Algorithms and Their Uses in Chemistry" in Reviews in Computational Chemistry, Vol. 10, pp. 1-73, K. B. Lipkowitz and D. B. Boyd, eds. (VCH Publishers, New York, 1997). Simulated annealing (Press et al., "Numerical Recipes in C: The Art of Scientific Computing", Cambridge University Press (Cambridge) 1992, Ch. 10), neural networks (E. Rich and K. Knight, "Artificial Intelligence", 2nd Edition (McGraw-Hill, New York, 1991, Ch. 18), standard gradient descent methods (Press et al., *supra,* Ch. 10), or other global or local optimization approaches (see discussion in Judson, *supra*) could also be used. Preferably, the correlation is found using a genetic algorithm approach as described in WO 01/01218.

Correlations may also be analyzed using analysis of variation (ANOVA) techniques to determine how much of the variation in the clinical data is explained by different subsets of the polymorphic sites in the SCYA8 gene. As described in WO 01/01218, ANOVA is used to test hypotheses about whether a response variable is caused by or correlated with one or more traits or

variables that can be measured (Fisher and vanBelle, *supra*, Ch. 10).

From the analyses described above, a mathematical model may be readily constructed by the skilled artisan that predicts clinical response as a function of SCYA8 genotype or haplotype content. Preferably, the model is validated in one or more follow-up clinical trials designed to test the model.

The identification of an association between a clinical response and a genotype or haplotype (or haplotype pair) for the SCYA8 gene may be the basis for designing a diagnostic method to determine those individuals who will or will not respond to the treatment, or alternatively, will respond at a lower level and thus may require more treatment, i.e., a greater dose of a drug. The diagnostic method may take one of several forms: for example, a direct DNA test (i.e., genotyping or haplotyping one or more of the polymorphic sites in the SCYA8 gene), a serological test, or a physical exam measurement. The only requirement is that there be a good correlation between the diagnostic test results and the underlying SCYA8 genotype or haplotype that is in turn correlated with the clinical response. In a preferred embodiment, this diagnostic method uses the predictive haplotyping method described above.

In another embodiment, the invention provides an isolated polynucleotide comprising a polymorphic variant of the SCYA8 gene or a fragment of the gene which contains at least one of the novel polymorphic sites described herein. The nucleotide sequence of a variant SCYA8 gene is identical to the reference genomic sequence for those portions of the gene examined, as described in the Examples below, except that it comprises a different nucleotide at one or more of the novel polymorphic sites PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11 and PS12. Similarly, the nucleotide sequence of a variant fragment of the SCYA8 gene is identical to the corresponding portion of the reference sequence except for having a different nucleotide at one or more of the novel polymorphic sites described herein. Thus, the invention specifically does not include polynucleotides comprising a nucleotide sequence identical to the reference sequence of the SCYA8 gene, which is defined by haplotype 5, (or other reported SCYA8 sequences) or to portions of the reference sequence (or other reported SCYA8 sequences), except for the haplotyping and genotyping oligonucleotides described above.

The location of a polymorphism in a variant SCYA8 gene or fragment is preferably identified by aligning its sequence against SEQ ID NO:1. The polymorphism is selected from the group consisting of thymine at PS1, adenine at PS2, cytosine at PS3, cytosine at PS4, cytosine at PS5, thymine at PS6, guanine at PS7, guanine at PS8, thymine at PS9, adenine at PS10, cytosine at PS11 and thymine at PS12. In a preferred embodiment, the polymorphic variant comprises a naturally-occurring isogene of the SCYA8 gene which is defined by any one of haplotypes 1- 4 and 6 - 11 shown in Table 5 below.

Polymorphic variants of the invention may be prepared by isolating a clone containing the SCYA8 gene from a human genomic library. The clone may be sequenced to determine the identity of the nucleotides at the novel polymorphic sites described herein. Any particular variant or fragment

thereof, that is claimed herein could be prepared from this clone by performing *in vitro* mutagenesis using procedures well-known in the art. Any particular SCYA8 variant or fragment thereof may also be prepared using synthetic or semi-synthetic methods known in the art.

SCYA8 isogenes, or fragments thereof, may be isolated using any method that allows separation of the two "copies" of the SCYA8 gene present in an individual, which, as readily understood by the skilled artisan, may be the same allele or different alleles. Separation methods include targeted *in vivo* cloning (TIVC) in yeast as described in WO 98/01573, U.S. Patent No. 5,866,404, and U.S. Patent No. 5,972,614. Another method, which is described in U.S. Patent No. 5,972,614, uses an allele specific oligonucleotide in combination with primer extension and exonuclease degradation to generate hemizygous DNA targets. Yet other methods are single molecule dilution (SMD) as described in Ruaño et al., *Proc. Natl. Acad. Sci.* 87:6296-6300, 1990; and allele specific PCR (Ruaño et al., 1989, *supra*; Ruaño et al., 1991, *supra*; Michalatos-Beloin et al., *supra*).

The invention also provides SCYA8 genome anthologies, which are collections of at least two SCYA8 isogenes found in a given population. The population may be any group of at least two individuals, including but not limited to a reference population, a population group, a family population, a clinical population, and a same gender population. A SCYA8 genome anthology may comprise individual SCYA8 isogenes stored in separate containers such as microtest tubes, separate wells of a microtitre plate and the like. Alternatively, two or more groups of the SCYA8 isogenes in the anthology may be stored in separate containers. Individual isogenes or groups of such isogenes in a genome anthology may be stored in any convenient and stable form, including but not limited to in buffered solutions, as DNA precipitates, freeze-dried preparations and the like. A preferred SCYA8 genome anthology of the invention comprises a set of isogenes defined by the haplotypes shown in Table 5 below. A SCYA8 genome anthology is useful in providing control nucleic acids for kits of the invention.

An isolated polynucleotide containing a polymorphic variant nucleotide sequence of the invention may be operably linked to one or more expression regulatory elements in a recombinant expression vector capable of being propagated and expressing the encoded SCYA8 protein in a prokaryotic or a eukaryotic host cell. Examples of expression regulatory elements which may be used include, but are not limited to, the lac system, operator and promoter regions of phage lambda, yeast promoters, and promoters derived from vaccinia virus, adenovirus, retroviruses, or SV40. Other regulatory elements include, but are not limited to, appropriate leader sequences, termination codons, polyadenylation signals, and other sequences required for the appropriate transcription and subsequent translation of the nucleic acid sequence in a given host cell. Of course, the correct combinations of expression regulatory elements will depend on the host system used. In addition, it is understood that the expression vector contains any additional elements necessary for its transfer to and subsequent replication in the host cell. Examples of such elements include, but are not limited to, origins of replication and selectable markers. Such expression vectors are commercially available or are readily

constructed using methods known to those in the art (e.g., F. Ausubel et al., 1987, in "Current Protocols in Molecular Biology", John Wiley and Sons, New York, New York). Host cells which may be used to express the variant SCYA8 sequences of the invention include, but are not limited to, eukaryotic and mammalian cells, such as animal, plant, insect and yeast cells, and prokaryotic cells, such as E. coli, or algal cells as known in the art. The recombinant expression vector may be introduced into the host cell using any method known to those in the art including, but not limited to, microinjection, electroporation, particle bombardment, transduction, and transfection using DEAE-dextran, lipofection, or calcium phosphate (see e.g., Sambrook et al. (1989) in "Molecular Cloning. A Laboratory Manual", Cold Spring Harbor Press, Plainview, New York). In a preferred aspect, eukaryotic expression vectors that function in eukaryotic cells, and preferably mammalian cells, are used. Non-limiting examples of such vectors include vaccinia virus vectors, adenovirus vectors, herpes virus vectors, and baculovirus transfer vectors. Preferred eukaryotic cell lines include COS cells, CHO cells, HeLa cells, NIH/3T3 cells, and embryonic stem cells (Thomson, J. A. et al., 1998 *Science* 282:1145-1147). Particularly preferred host cells are mammalian cells.

As will be readily recognized by the skilled artisan, expression of polymorphic variants of the SCYA8 gene will produce SCYA8 mRNAs varying from each other at any polymorphic site retained in the spliced and processed mRNA molecules. These mRNAs can be used for the preparation of a SCYA8 cDNA comprising a nucleotide sequence which is a polymorphic variant of the SCYA8 reference coding sequence shown in Figure 2. Thus, the invention also provides SCYA8 mRNAs and corresponding cDNAs which comprise a nucleotide sequence that is identical to SEQ ID NO:2 (Fig. 2), or its corresponding RNA sequence, for those regions of SEQ ID NO:2 that correspond to the examined portions of the SCYA8 gene (as described in the Examples below), except for having one or more polymorphisms selected from the group consisting of thymine at a position corresponding to nucleotide 117, adenine at a position corresponding to nucleotide 201, cytosine at a position corresponding to nucleotide 235 and thymine at a position corresponding to nucleotide 250. A particularly preferred polymorphic cDNA variant comprises the coding sequence of a SCYA8 isogene defined by any one of haplotypes 1-4, 6, 8, 9 and 11. Fragments of these variant mRNAs and cDNAs are included in the scope of the invention, provided they contain one or more of the novel polymorphisms described herein. The invention specifically excludes polynucleotides identical to previously identified and characterized SCYA8 mRNAs, cDNAs or fragments thereof. Polynucleotides comprising a variant SCYA8 RNA or DNA sequence may be isolated from a biological sample using well-known molecular biological procedures or may be chemically synthesized.

As used herein, a polymorphic variant of a SCYA8 gene, mRNA or cDNA fragment comprises at least one novel polymorphism identified herein and has a length of at least 10 nucleotides and may range up to the full length of the gene. Preferably, such fragments are between 100 and 3000 nucleotides in length, and more preferably between 200 and 2000 nucleotides in length, and most

preferably between 500 and 1000 nucleotides in length.

In describing the SCYA8 polymorphic sites identified herein, reference is made to the sense strand of the gene for convenience. However, as recognized by the skilled artisan, nucleic acid molecules containing the SCYA8 gene or cDNA may be complementary double stranded molecules and thus reference to a particular site on the sense strand refers as well to the corresponding site on the complementary antisense strand. Thus, reference may be made to the same polymorphic site on either strand and an oligonucleotide may be designed to hybridize specifically to either strand at a target region containing the polymorphic site. Thus, the invention also includes single-stranded polynucleotides which are complementary to the sense strand of the SCYA8 genomic, mRNA and cDNA variants described herein.

Polynucleotides comprising a polymorphic gene variant or fragment of the invention may be useful for therapeutic purposes. For example, where a patient could benefit from expression, or increased expression, of a particular SCYA8 protein isoform, an expression vector encoding the isoform may be administered to the patient. The patient may be one who lacks the SCYA8 isogene encoding that isoform or may already have at least one copy of that isogene.

In other situations, it may be desirable to decrease or block expression of a particular SCYA8 isogene. Expression of a SCYA8 isogene may be turned off by transforming a targeted organ, tissue or cell population with an expression vector that expresses high levels of untranslatable mRNA or antisense RNA for the isogene or fragment thereof. Alternatively, oligonucleotides directed against the regulatory regions (e.g., promoter, introns, enhancers, 3′ untranslated region) of the isogene may block transcription. Oligonucleotides targeting the transcription initiation site, e.g., between positions −10 and +10 from the start site are preferred. Similarly, inhibition of transcription can be achieved using oligonucleotides that base-pair with region(s) of the isogene DNA to form triplex DNA (see e.g., Gee et al. in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing Co., Mt. Kisco, N.Y., 1994). Antisense oligonucleotides may also be designed to block translation of SCYA8 mRNA transcribed from a particular isogene. It is also contemplated that ribozymes may be designed that can catalyze the specific cleavage of SCYA8 mRNA transcribed from a particular isogene.

The untranslated mRNA, antisense RNA or antisense oligonucleotides may be delivered to a target cell or tissue by expression from a vector introduced into the cell or tissue *in vivo* or *ex vivo*. Alternatively, such molecules may be formulated as a pharmaceutical composition for administration to the patient. Oligoribonucleotides and/or oligodeoxynucleotides intended for use as antisense oligonucleotides may be modified to increase stability and half-life. Possible modifications include, but are not limited to phosphorothioate or 2′ O-methyl linkages, and the inclusion of nontraditional bases such as inosine and queosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytosine, guanine, thymine, and uracil which are not as easily recognized by endogenous nucleases.

The invention also provides an isolated polypeptide comprising a polymorphic variant of (a) the reference SCYA8 amino acid sequence shown in Figure 3 or (b) a fragment of this reference sequence. The location of a variant amino acid in a SCYA8 polypeptide or fragment of the invention is identified by aligning its sequence against SEQ ID NO:3 (Fig. 3). A SCYA8 protein variant of the invention comprises an amino acid sequence identical to SEQ ID NO:3 for those regions of SEQ ID NO:3 that are encoded by examined portions of the SCYA8 gene (as described in the Examples below), except for having one or more variant amino acids selected from the group consisting ofglutamine at a position corresponding to amino acid position 79 and phenylalanine at a position corresponding to amino acid position 84. Thus, a SCYA8 fragment of the invention, also referred to herein as a SCYA8 peptide variant, is any fragment of a SCYA8 protein variant that contains one or more of the amino acid variations shown in Table 2. The invention specifically excludes amino acid sequences identical to those previously identified for SCYA8, including SEQ ID NO:3, and previously described fragments thereof. SCYA8 protein variants included within the invention comprise all amino acid sequences based on SEQ ID NO:3 and having the combination of amino acid variations described in Table 2 below. In preferred embodiments, a SCYA8 protein variant of the invention is encoded by an isogene defined by one of the observed haplotypes, 1-4, 6, 8, 9 and 11, shown in Table 5.

Table 2. Novel Polymorphic Variants of SCYA8

Polymorphic Variant Number	Amino Acid Position and Identities	
	79	84
1	K	F
2	Q	V
3	Q	F

A SCYA8 peptide variant of the invention is at least 6 amino acids in length and is preferably any number between 6 and 30 amino acids long, more preferably between 10 and 25, and most preferably between 15 and 20 amino acids long. Such SCYA8 peptide variants may be useful as antigens to generate antibodies specific for one of the above SCYA8 isoforms. In addition, the SCYA8 peptide variants may be useful in drug screening assays.

A SCYA8 variant protein or peptide of the invention may be prepared by chemical synthesis or by expressing an appropriate variant SCYA8 genomic or cDNA sequence described above. Alternatively, the SCYA8 protein variant may be isolated from a biological sample of an individual having a SCYA8 isogene which encodes the variant protein. Where the sample contains two different SCYA8 isoforms (i.e., the individual has different SCYA8 isogenes), a particular SCYA8 isoform of the invention can be isolated by immunoaffinity chromatography using an antibody which specifically binds to that particular SCYA8 isoform but does not bind to the other SCYA8 isoform.

The expressed or isolated SCYA8 protein or peptide may be detected by methods known in the

art, including Coomassie blue staining, silver staining, and Western blot analysis using antibodies specific for the isoform of the SCYA8 protein or peptide as discussed further below. SCYA8 variant proteins and peptides can be purified by standard protein purification procedures known in the art, including differential precipitation, molecular sieve chromatography, ion-exchange chromatography, isoelectric focusing, gel electrophoresis, affinity and immunoaffinity chromatography and the like. (Ausubel et. al., 1987, In Current Protocols in Molecular Biology John Wiley and Sons, New York, New York). In the case of immunoaffinity chromatography, antibodies specific for a particular polymorphic variant may be used.

A polymorphic variant SCYA8 gene of the invention may also be fused in frame with a heterologous sequence to encode a chimeric SCYA8 protein. The non-SCYA8 portion of the chimeric protein may be recognized by a commercially available antibody. In addition, the chimeric protein may also be engineered to contain a cleavage site located between the SCYA8 and non-SCYA8 portions so that the SCYA8 protein may be cleaved and purified away from the non-SCYA8 portion.

An additional embodiment of the invention relates to using a novel SCYA8 protein isoform, or a fragment thereof, in any of a variety of drug screening assays. Such screening assays may be performed to identify agents that bind specifically to all known SCYA8 protein isoforms or to only a subset of one or more of these isoforms. The agents may be from chemical compound libraries, peptide libraries and the like. The SCYA8 protein or peptide variant may be free in solution or affixed to a solid support. In one embodiment, high throughput screening of compounds for binding to a SCYA8 variant may be accomplished using the method described in PCT application WO84/03565, in which large numbers of test compounds are synthesized on a solid substrate, such as plastic pins or some other surface, contacted with the SCYA8 protein(s) of interest and then washed. Bound SCYA8 protein(s) are then detected using methods well-known in the art.

In another embodiment, a novel SCYA8 protein isoform may be used in assays to measure the binding affinities of one or more candidate drugs targeting the SCYA8 protein.

In yet another embodiment, when a particular SCYA8 haplotype or group of SCYA8 haplotypes encodes a SCYA8 protein variant with an amino acid sequence distinct from that of SCYA8 protein isoforms encoded by other SCYA8 haplotypes, then detection of that particular SCYA8 haplotype or group of SCYA8 haplotypes may be accomplished by detecting expression of the encoded SCYA8 protein variant using any of the methods described herein or otherwise commonly known to the skilled artisan.

In another embodiment, the invention provides antibodies specific for and immunoreactive with one or more of the novel SCYA8 variant proteins described herein. The antibodies may be either monoclonal or polyclonal in origin. The SCYA8 protein or peptide variant used to generate the antibodies may be from natural or recombinant sources or produced by chemical synthesis using synthesis techniques known in the art. If the SCYA8 protein variant is of insufficient size to be antigenic, it may be conjugated, complexed, or otherwise covalently linked to a carrier molecule to

enhance the antigenicity of the peptide. Examples of carrier molecules, include, but are not limited to, albumins (e.g., human, bovine, fish, ovine), and keyhole limpet hemocyanin (Basic and Clinical Immunology, 1991, Eds. D.P. Stites, and A.I. Terr, Appleton and Lange, Norwalk Connecticut, San Mateo, California).

In one embodiment, an antibody specifically immunoreactive with one of the novel protein isoforms described herein is administered to an individual to neutralize activity of the SCYA8 isoform expressed by that individual. The antibody may be formulated as a pharmaceutical composition which includes a pharmaceutically acceptable carrier.

Antibodies specific for and immunoreactive with one of the novel protein isoforms described herein may be used to immunoprecipitate the SCYA8 protein variant from solution as well as react with SCYA8 protein isoforms on Western or immunoblots of polyacrylamide gels on membrane supports or substrates. In another preferred embodiment, the antibodies will detect SCYA8 protein isoforms in paraffin or frozen tissue sections, or in cells which have been fixed or unfixed and prepared on slides, coverslips, or the like, for use in immunocytochemical, immunohistochemical, and immunofluorescence techniques.

In another embodiment, an antibody specifically immunoreactive with one of the novel SCYA8 protein variants described herein is used in immunoassays to detect this variant in biological samples. In this method, an antibody of the present invention is contacted with a biological sample and the formation of a complex between the SCYA8 protein variant and the antibody is detected. As described, suitable immunoassays include radioimmunoassay, Western blot assay, immunofluorescent assay, enzyme linked immunoassay (ELISA), chemiluminescent assay, immunohistochemical assay, immunocytochemical assay, and the like (see, e.g., Principles and Practice of Immunoassay, 1991, Eds. Christopher P. Price and David J. Neoman, Stockton Press, New York, New York; Current Protocols in Molecular Biology, 1987, Eds. Ausubel et al., John Wiley and Sons, New York, New York). Standard techniques known in the art for ELISA are described in Methods in Immunodiagnosis, 2nd Ed., Eds. Rose and Bigazzi, John Wiley and Sons, New York 1980; and Campbell et al., 1984, Methods in Immunology, W.A. Benjamin, Inc.). Such assays may be direct, indirect, competitive, or noncompetitive as described in the art (see, e.g., Principles and Practice of Immunoassay, 1991, Eds. Christopher P. Price and David J. Neoman, Stockton Pres, NY, NY; and Oellirich, M., 1984, J. Clin. Chem. Clin. Biochem., 22:895-904). Proteins may be isolated from test specimens and biological samples by conventional methods, as described in Current Protocols in Molecular Biology, supra.

Exemplary antibody molecules for use in the detection and therapy methods of the present invention are intact immunoglobulin molecules, substantially intact immunoglobulin molecules, or those portions of immunoglobulin molecules that contain the antigen binding site. Polyclonal or monoclonal antibodies may be produced by methods conventionally known in the art (e.g., Kohler and Milstein, 1975, Nature, 256:495-497; Campbell Monoclonal Antibody Technology, the Production and Characterization of Rodent and Human Hybridomas, 1985, In: Laboratory Techniques in Biochemistry

and Molecular Biology, Eds. Burdon et al., Volume 13, Elsevier Science Publishers, Amsterdam). The antibodies or antigen binding fragments thereof may also be produced by genetic engineering. The technology for expression of both heavy and light chain genes in E. coli is the subject of PCT patent applications, publication number WO 901443, WO 901443 and WO 9014424 and in Huse et al., 1989, Science, 246:1275-1281. The antibodies may also be humanized (e.g., Queen, C. et al. 1989 Proc. Natl. Acad. Sci.USA 86;10029).

Effect(s) of the polymorphisms identified herein on expression of SCYA8 may be investigated by preparing recombinant cells and/or nonhuman recombinant organisms, preferably recombinant animals, containing a polymorphic variant of the SCYA8 gene. As used herein, "expression" includes but is not limited to one or more of the following: transcription of the gene into precursor mRNA; splicing and other processing of the precursor mRNA to produce mature mRNA; mRNA stability; translation of the mature mRNA into SCYA8 protein (including codon usage and tRNA availability); and glycosylation and/or other modifications of the translation product, if required for proper expression and function.

To prepare a recombinant cell of the invention, the desired SCYA8 isogene may be introduced into the cell in a vector such that the isogene remains extrachromosomal. In such a situation, the gene will be expressed by the cell from the extrachromosomal location. In a preferred embodiment, the SCYA8 isogene is introduced into a cell in such a way that it recombines with the endogenous SCYA8 gene present in the cell. Such recombination requires the occurrence of a double recombination event, thereby resulting in the desired SCYA8 gene polymorphism. Vectors for the introduction of genes both for recombination and for extrachromosomal maintenance are known in the art, and any suitable vector or vector construct may be used in the invention. Methods such as electroporation, particle bombardment, calcium phosphate co-precipitation and viral transduction for introducing DNA into cells are known in the art; therefore, the choice of method may lie with the competence and preference of the skilled practitioner. Examples of cells into which the SCYA8 isogene may be introduced include, but are not limited to, continuous culture cells, such as COS, NIH/3T3, and primary or culture cells of the relevant tissue type, i.e., they express the SCYA8 isogene. Such recombinant cells can be used to compare the biological activities of the different protein variants.

Recombinant nonhuman organisms, i.e., transgenic animals, expressing a variant SCYA8 gene are prepared using standard procedures known in the art. Preferably, a construct comprising the variant gene is introduced into a nonhuman animal or an ancestor of the animal at an embryonic stage, i.e., the one-cell stage, or generally not later than about the eight-cell stage. Transgenic animals carrying the constructs of the invention can be made by several methods known to those having skill in the art. One method involves transfecting into the embryo a retrovirus constructed to contain one or more insulator elements, a gene or genes of interest, and other components known to those skilled in the art to provide a complete shuttle vector harboring the insulated gene(s) as a transgene, see e.g., U.S. Patent No. 5,610,053. Another method involves directly injecting a transgene into the embryo. A

third method involves the use of embryonic stem cells. Examples of animals into which the SCYA8 isogenes may be introduced include, but are not limited to, mice, rats, other rodents, and nonhuman primates (see "The Introduction of Foreign Genes into Mice" and the cited references therein, In: Recombinant DNA, Eds. J.D. Watson, M. Gilman, J. Witkowski, and M. Zoller; W.H. Freeman and Company, New York, pages 254-272). Transgenic animals stably expressing a human SCYA8 isogene and producing the encoded human SCYA8 protein can be used as biological models for studying diseases related to abnormal SCYA8 expression and/or activity, and for screening and assaying various candidate drugs, compounds, and treatment regimens to reduce the symptoms or effects of these diseases.

An additional embodiment of the invention relates to pharmaceutical compositions for treating disorders affected by expression or function of a novel SCYA8 isogene described herein. The pharmaceutical composition may comprise any of the following active ingredients: a polynucleotide comprising one of these novel SCYA8 isogenes; an antisense oligonucleotide directed against one of the novel SCYA8 isogenes, a polynucleotide encoding such an antisense oligonucleotide, or another compound which inhibits expression of a novel SCYA8 isogene described herein. Preferably, the composition contains the active ingredient in a therapeutically effective amount. By therapeutically effective amount is meant that one or more of the symptoms relating to disorders affected by expression or function of a novel SCYA8 isogene is reduced and/or eliminated. The composition also comprises a pharmaceutically acceptable carrier, examples of which include, but are not limited to, saline, buffered saline, dextrose, and water. Those skilled in the art may employ a formulation most suitable for the active ingredient, whether it is a polynucleotide, oligonucleotide, protein, peptide or small molecule antagonist. The pharmaceutical composition may be administered alone or in combination with at least one other agent, such as a stabilizing compound. Administration of the pharmaceutical composition may be by any number of routes including, but not limited to oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, intradermal, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal. Further details on techniques for formulation and administration may be found in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing Co., Easton, PA).

For any composition, determination of the therapeutically effective dose of active ingredient and/or the appropriate route of administration is well within the capability of those skilled in the art. For example, the dose can be estimated initially either in cell culture assays or in animal models. The animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans. The exact dosage will be determined by the practitioner, in light of factors relating to the patient requiring treatment, including but not limited to severity of the disease state, general health, age, weight and gender of the patient, diet, time and frequency of administration, other drugs being taken by the patient, and tolerance/response to the treatment.

Any or all analytical and mathematical operations involved in practicing the methods of the present invention may be implemented by a computer. In addition, the computer may execute a program that generates views (or screens) displayed on a display device and with which the user can interact to view and analyze large amounts of information relating to the SCYA8 gene and its genomic variation, including chromosome location, gene structure, and gene family, gene expression data, polymorphism data, genetic sequence data, and clinical data population data (e.g., data on ethnogeographic origin, clinical responses, genotypes, and haplotypes for one or more populations). The SCYA8 polymorphism data described herein may be stored as part of a relational database (e.g., an instance of an Oracle database or a set of ASCII flat files). These polymorphism data may be stored on the computer's hard drive or may, for example, be stored on a CD-ROM or on one or more other storage devices accessible by the computer. For example, the data may be stored on one or more databases in communication with the computer via a network.

Preferred embodiments of the invention are described in the following examples. Other embodiments within the scope of the claims herein will be apparent to one skilled in the art from consideration of the specification or practice of the invention as disclosed herein. It is intended that the specification, together with the examples, be considered exemplary only, with the scope and spirit of the invention being indicated by the claims which follow the examples.

EXAMPLES

The Examples herein are meant to exemplify the various aspects of carrying out the invention and are not intended to limit the scope of the invention in any way. The Examples do not include detailed descriptions for conventional methods employed, such as in the performance of genomic DNA isolation, PCR and sequencing procedures. Such methods are well-known to those skilled in the art and are described in numerous publications, for example, Sambrook, Fritsch, and Maniatis, "Molecular Cloning: A Laboratory Manual", 2nd Edition, Cold Spring Harbor Laboratory Press, USA, (1989).

EXAMPLE 1

This example illustrates examination of various regions of the SCYA8 gene for polymorphic sites.

Amplification of Target Regions

The following target regions of the SCYA8 gene were amplified using PCR primer pairs. The primers used for each region are represented below by providing the nucleotide positions of their initial and final nucleotides, which correspond to positions in SEQ ID NO:1 (Figure 1).

PCR Primer Pairs

Fragment	Forward Primer	Reverse Primer		PCR Product
Fragment 1	3321 - 3344	complement of	3984 - 3962	664 nt
Fragment 2	3581 - 3602	complement of	4279 - 4254	699 nt
Fragment 3	3788 - 3816	complement of	4333 - 4313	546 nt
Fragment 4	4573 - 4596	complement of	5115 - 5092	543 nt
Fragment 5	5201 - 5226	complement of	5801 - 5779	601 nt

These primer pairs were used in PCR reactions containing genomic DNA isolated from immortalized cell lines for each member of the Index Repository. The PCR reactions were carried out under the following conditions:

Reaction volume	= 10 µl
10 x Advantage 2 Polymerase reaction buffer (Clontech)	= 1 µl
100 ng of human genomic DNA	= 1 µl
10 mM dNTP	= 0.4 µl
Advantage 2 Polymerase enzyme mix (Clontech)	= 0.2 µl
Forward Primer (10 µM)	= 0.4 µl
Reverse Primer (10 µM)	= 0.4 µl
Water	= 6.6µl

Amplification profile:
97°C - 2 min. 1 cycle

97°C - 15 sec.
70°C - 45 sec. } 10 cycles
72°C - 45 sec.

97°C - 15 sec.
64°C - 45 sec. } 35 cycles
72°C - 45 sec.

Sequencing of PCR Products

The PCR products were purified using a Whatman/Polyfiltronics 100 µl 384 well unifilter plate essentially according to the manufacturers protocol. The purified DNA was eluted in 50 µl of distilled water. Sequencing reactions were set up using Applied Biosystems Big Dye Terminator chemistry essentially according to the manufacturers protocol. The purified PCR products were sequenced in both directions using the primer sets described previously or those represented below by the nucleotide positions of their initial and final nucleotides, which correspond to positions in SEQ ID NO:1 (Figure 1). Reaction products were purified by isopropanol precipitation, and run on an Applied Biosystems 3700 DNA Analyzer.

Sequencing Primer Pairs

Fragment	Forward Primer	Reverse Primer	
Fragment 1	3357 - 3376	complement of	3902 - 3882
Fragment 2	3652 - 3672	complement of	4172 - 4151
Fragment 3	3886 - 3907	complement of	4292 - 4273
Fragment 4	4655 - 4674	complement of	5074 - 5055
Fragment 5	5232 - 5250	complement of	5712 - 5693

Analysis of Sequences for Polymorphic Sites

Sequence information for a minimum of 80 humans was analyzed for the presence of polymorphisms using the Polyphred program (Nickerson et al., *Nucleic Acids Res.* 14:2745-2751, 1997). The presence of a polymorphism was confirmed on both strands. The polymorphisms and their locations in the SCYA8 reference genomic sequence (SEQ ID NO:1) are listed in Table 3 below.

Table 3. Polymorphic Sites Identified in the SCYA8 Gene

Polymorphic Site Number	PolyId[a]	Nucleotide Position	Reference Allele	Variant Allele	CDS Variant Position	AA Variant
PS1	1097869	3499	C	T		
PS2	1097863	3681	G	A		
PS3	1097853	3901	T	C		
PS4	1097851	3953	A	C		
PS5	1097849	3961	G	C		
PS6	1097845	3996	C	T		
PS7	1097843	4019	A	G		
PS8	1097829	4803	C	G		
PS9	1097827	4848	C	T	117	S39S
PS10	1097825	4932	C	A	201	I67I
PS11	1097817	5381	A	C	235	K79Q
PS12	1097815	5396	G	T	250	V84F

[a]PolyId is a unique identifier assigned to each PS by Genaissance Pharmaceuticals, Inc.

EXAMPLE 2

This example illustrates analysis of the SCYA8 polymorphisms identified in the Index Repository for human genotypes and haplotypes.

The different genotypes containing these polymorphisms that were observed in unrelated members of the reference population are shown in Table 4 below, with the haplotype pair indicating the combination of haplotypes determined for the individual using the haplotype derivation protocol described below. In Table 4, homozygous positions are indicated by one nucleotide and heterozygous positions are indicated by two nucleotides. Missing nucleotides in any given genotype in Table 4 were inferred based on linkage disequilibrium and/or Mendelian inheritance.

Table 4 (Part1). Genotypes and Haplotype Pairs Observed for SCYA8 Gene

Genotype Number	PS1	PS2	PS3	PS4	PS5	PS6	PS7	PS8	PS9	PS10	HAP	Pair
1	C	G	T	A	G	C	A	C	C	C	5	5
2	T	G	T	A	G	C	A	C	C	C	10	10
3	C	G	T	A	G	C	A	C/G	C	C	5	7
4	C	G	T	A	G	C	A/G	C	C	C	5	8
5	C	A	T	A	G	C	A	C	C/T	C	1	2
6	C	G	T	A	G/C	C	A	C	C	C	5	4
7	C/T	G	T	A	G	C	A	C	C	C	5	11
8	C	G	T/C	A/C	G	C	A/G	C	C	C	5	3
9	C	G	T	A	G	C/T	A	C	C	C/A	5	9
10	C/T	G	T	A	G	C	A	C	C	C	5	10
11	C	G/A	T	A	G	C	A	C	C	C	5	1
12	C	G	T	A	G	C	A	C	C	C	5	6

Table 4(Part2). Genotypes and Haplotype Pairs Observed for SCYA8 Gene

Genotype Number	PS11	PS12	HAP	Pair
1	A	G	5	5
2	A	G	10	10
3	A	G	5	7
4	A/C	G	5	8
5	C	G	1	2
6	A/C	G	5	4
7	A/C	G	5	11
8	A/C	G	5	3
9	A	G	5	9
10	A	G	5	10
11	A/C	G	5	1
12	A	G/T	5	6

The haplotype pairs shown in Table 4 were estimated from the unphased genotypes using a computer-implemented extension of Clark's algorithm (Clark, A.G. 1990 *Mol Bio Evol* 7, 111-122) for assigning haplotypes to unrelated individuals in a population sample, as described in PCT/US01/12831, filed April 18, 2001. In this method, haplotypes are assigned directly from individuals who are homozygous at all sites or heterozygous at no more than one of the variable sites. This list of haplotypes is then used to deconvolute the unphased genotypes in the remaining (multiply heterozygous) individuals. In the present analysis, the list of haplotypes was augmented with haplotypes obtained from two families (one three-generation Caucasian family and one two-generation African-American family).

By following this protocol, it was determined that the Index Repository examined herein and, by extension, the general population contains the 11 human SCYA8 haplotypes shown in Table 5 below.

An SCYA8 isogene defined by a full-haplotype shown in Table 5 below comprises the regions of the SEQ ID NOS indicated in Table 5, with their corresponding set of polymorphic locations and

identities, which are also set forth in Table 5.

Table 5. Haplotypes of the SCYA8 Gene

Haplotype Number[a]											PS	PS	SEQ ID	Regions
1	2	3	4	5	6	7	8	9	10	11	No.[b]	Position[c]	No.[d]	Examined[e]
C	C	C	C	C	C	C	C	C	T	T	1	3499/30	1/64	3321-4333
A	A	G	G	G	G	G	G	G	G	G	2	3681/150	1/64	3321-4333
T	T	C	T	T	T	T	T	T	T	T	3	3901/270	1/64	3321-4333
A	A	C	A	A	A	A	A	A	A	A	4	3953/390	1/64	3321-4333
G	G	G	C	G	G	G	G	G	G	G	5	3961/510	1/64	3321-4333
C	C	C	C	C	C	C	C	T	C	C	6	3996/630	1/64	3321-4333
A	A	G	A	A	A	A	G	A	A	A	7	4019/750	1/64	3321-4333
C	C	C	C	C	C	G	C	C	C	C	8	4803/870	1/64	4573-5115
C	T	C	C	C	C	C	C	C	C	C	9	4848/990	1/64	4573-5115
C	C	C	C	C	C	C	C	A	C	C	10	4932/1110	1/64	4573-5115
C	C	C	C	A	A	A	C	A	A	C	11	5381/1230	1/64	5201-5801
G	G	G	G	G	T	G	G	G	G	G	12	5396/1350	1/64	5201-5801

[a]Alleles for SCYA8 haplotypes are presented 5′ to 3′ in each column;
[b]PS = polymorphic site;
[c]Position of PS within the indicated SEQ ID NO, with the 1st position number referring to the first SEQ ID NO and the 2nd position number referring to the 2nd SEQ ID NO;
[d]1st SEQ ID NO refers to Figure 1, with the two alternative allelic variants of each polymorphic site indicated by the appropriate nucleotide symbol; 2nd SEQ ID NO is a modified version of the 1st SEQ ID NO that comprises the context sequence of each polymorphic site, PS1-PS12, to facilitate electronic searching of the haplotypes;
[e]Region examined represents the nucleotide positions defining the start and stop positions within the 1st SEQ ID NO of the sequenced region.

SEQ ID NO:1 refers to Figure 1, with the two alternative allelic variants of each polymorphic site indicated by the appropriate nucleotide symbol. SEQ ID NO:64 is a modified version of SEQ ID NO:1 that shows the context sequence of each of PS1-PS12 in a uniform format to facilitate electronic searching of the SCYA8 haplotypes. For each polymorphic site, SEQ ID NO:64 contains a block of 60 bases of the nucleotide sequence encompassing the centrally-located polymorphic site at the 30th position, followed by 60 bases of unspecified sequence to represent that each polymorphic site is separated by genomic sequence whose composition is defined elsewhere herein.

Table 6 below shows the percent of chromosomes characterized by a given SCYA8 haplotype for all unrelated individuals in the Index Repository for which haplotype data was obtained. The percent of these unrelated individuals who have a given SCYA8 haplotype pair is shown in Table 7. In Tables 6 and 7, the "Total" column shows this frequency data for all of these unrelated individuals, while the other columns show the frequency data for these unrelated individuals categorized according to their self-identified ethnogeographic origin. Abbreviations used in Tables 6 and 7 are AF = African Descent, AS = Asian, CA = Caucasian, HL = Hispanic-Latino, and AM = Native American.

Table 6. Frequency of Observed SCYA8 Haplotypes In Unrelated Individuals

HAP No.	HAP ID	Total	CA	AF	AS	HL	AM
1	1100781	13.41	7.14	5.0	25.0	13.89	33.33
2	1100788	0.61	0.0	0.0	2.5	0.0	0.0
3	1100783	1.22	0.0	5.0	0.0	0.0	0.0
4	1100787	0.61	2.38	0.0	0.0	0.0	0.0
5	1100779	60.98	54.76	72.5	55.0	63.89	50.0
6	1100789	0.61	0.0	2.5	0.0	0.0	0.0
7	1100790	0.61	0.0	0.0	0.0	2.78	0.0
8	1100786	0.61	0.0	0.0	0.0	0.0	16.67
9	1100782	1.22	0.0	5.0	0.0	0.0	0.0
10	1100780	18.9	33.33	10.0	17.5	16.67	0.0
11	1100785	1.22	2.38	0.0	0.0	2.78	0.0

Table 7. Frequency of Observed SCYA8 Haplotype Pairs In Unrelated Individuals

HAP1	HAP2	Total	CA	AF	AS	HL	AM
5	5	29.27	23.81	45.0	20.0	33.33	0.0
10	10	6.1	14.29	0.0	5.0	5.56	0.0
5	7	1.22	0.0	0.0	0.0	5.56	0.0
5	8	1.22	0.0	0.0	0.0	0.0	33.33
1	2	1.22	0.0	0.0	5.0	0.0	0.0
5	4	1.22	4.76	0.0	0.0	0.0	0.0
5	11	2.44	4.76	0.0	0.0	5.56	0.0
5	3	2.44	0.0	10.0	0.0	0.0	0.0
5	9	2.44	0.0	10.0	0.0	0.0	0.0
5	10	25.61	38.1	20.0	25.0	22.22	0.0
5	1	25.61	14.29	10.0	45.0	27.78	66.67
5	6	1.22	0.0	5.0	0.0	0.0	0.0

The size and composition of the Index Repository were chosen to represent the genetic diversity across and within four major population groups comprising the general United States population. For example, as described in Table 1 above, this repository contains approximately equal sample sizes of African-descent, Asian-American, European-American, and Hispanic-Latino population groups. Almost all individuals representing each group had all four grandparents with the same ethnogeographic background. The number of unrelated individuals in the Index Repository provides a sample size that is sufficient to detect SNPs and haplotypes that occur in the general population with high statistical certainty. For instance, a haplotype that occurs with a frequency of 5% in the general population has a probability higher than 99.9% of being observed in a sample of 80 individuals from the general population. Similarly, a haplotype that occurs with a frequency of 10% in a specific population group has a 99% probability of being observed in a sample of 20 individuals from that population group. In addition, the size and composition of the Index Repository means that the relative frequencies determined therein for the haplotypes and haplotype pairs of the SCYA8 gene are likely to be similar to the relative frequencies of these SCYA8 haplotypes and haplotype pairs in the general U.S. population and in the four population groups represented in the Index Repository. The

genetic diversity observed for the three Native Americans is presented because it is of scientific interest, but due to the small sample size it lacks statistical significance.

In view of the above, it will be seen that the several advantages of the invention are achieved and other advantageous results attained.

As various changes could be made in the above methods and compositions without departing from the scope of the invention, it is intended that all matter contained in the above description and shown in the accompanying drawings shall be interpreted as illustrative and not in a limiting sense.

All references cited in this specification, including patents and patent applications, are hereby incorporated in their entirety by reference. The discussion of references herein is intended merely to summarize the assertions made by their authors and no admission is made that any reference constitutes prior art. Applicants reserve the right to challenge the accuracy and pertinency of the cited references.

What is Claimed is:

1. A method for haplotyping the small inducible cytokine subfamily A (Cys-Cys), member 8 (monocyte chemotactic protein 2) (SCYA8) gene of an individual, which comprises determining which of the SCYA8 haplotypes shown in the table immediately below defines one copy of the individual's SCYA8 gene, wherein each of the SCYA8 haplotypes comprises a sequence of polymorphisms whose positions and identities are set forth in the table immediately below:

Haplotype Number[a]											PS	PS
1	2	3	4	5	6	7	8	9	10	11	No.[b]	Position[c]
C	C	C	C	C	C	C	C	C	T	T	1	3499
A	A	G	G	G	G	G	G	G	G	G	2	3681
T	T	C	T	T	T	T	T	T	T	T	3	3901
A	A	C	A	A	A	A	A	A	A	A	4	3953
G	G	G	C	G	G	G	G	G	G	G	5	3961
C	C	C	C	C	C	C	C	T	C	C	6	3996
A	A	G	A	A	A	A	G	A	A	A	7	4019
C	C	C	C	C	C	G	C	C	C	C	8	4803
C	T	C	C	C	C	C	C	C	C	C	9	4848
C	C	C	C	C	C	C	C	A	C	C	10	4932
C	C	C	C	A	A	A	C	A	A	C	11	5381
G	G	G	G	G	T	G	G	G	G	G	12	5396

[a]Alleles for haplotypes are presented 5′ to 3′ in each column;
[b]PS = polymorphic site;
[c]Position of PS within SEQ ID NO:1.

2. The method of claim 1, wherein the determining step comprises identifying the phased sequence of nucleotides present at each of PS1-PS12 on at least one copy of the individual's SCYA8 gene.

3. A method for haplotyping the small inducible cytokine subfamily A (Cys-Cys), member 8 (monocyte chemotactic protein 2) (SCYA8) gene of an individual, which comprises determining which of the SCYA8 haplotype pairs shown in the table immediately below defines both copies of the individual's SCYA8 gene, wherein each of the SCYA8 haplotype pairs consists of first and second haplotypes which comprise first and second sequences of polymorphisms whose positions and identities are set forth in the table immediately below:

Haplotype Pair[a] (Part 1)								PS	PS
5/5	10/10	5/7	5/8	1/2	5/4	5/11	5/3	No.[b]	Position[c]
C/C	T/T	C/C	C/C	C/C	C/C	C/T	C/C	1	3499
G/G	G/G	G/G	G/G	A/A	G/G	G/G	G/G	2	3681
T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/C	3	3901
A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/C	4	3953
G/G	G/G	G/G	G/G	G/G	G/C	G/G	G/G	5	3961
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	6	3996
A/A	A/A	A/A	A/G	A/A	A/A	A/A	A/G	7	4019
C/C	C/C	C/G	C/C	C/C	C/C	C/C	C/C	8	4803
C/C	C/C	C/C	C/C	C/T	C/C	C/C	C/C	9	4848
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	10	4932
A/A	A/A	A/A	A/C	C/C	A/C	A/C	A/C	11	5381
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	12	5396

Haplotype Pair[a] (Part 2)				PS	PS
5/9	5/10	5/1	5/6	No.[b]	Position[c]
C/C	C/T	C/C	C/C	1	3499
G/G	G/G	G/A	G/G	2	3681
T/T	T/T	T/T	T/T	3	3901
A/A	A/A	A/A	A/A	4	3953
G/G	G/G	G/G	G/G	5	3961
C/T	C/C	C/C	C/C	6	3996
A/A	A/A	A/A	A/A	7	4019
C/C	C/C	C/C	C/C	8	4803
C/C	C/C	C/C	C/C	9	4848
C/A	C/C	C/C	C/C	10	4932
A/A	A/A	A/C	A/A	11	5381
G/G	G/G	G/G	G/T	12	5396

[a]Haplotype pairs are represented as 1^{st} haplotype/2^{nd} haplotype; with alleles of each haplotype shown 5′ to 3′ as 1^{st} polymorphism/2^{nd} polymorphism in each column;
[b]PS = polymorphic site;
[c]Position of PS in SEQ ID NO:1.

4. The method of claim 3, wherein the determining step comprises identifying the phased sequence of nucleotides present at each of PS1-PS12 on both copies of the individual's SCYA8 gene.

5. A method for genotyping the small inducible cytokine subfamily A (Cys-Cys), member 8 (monocyte chemotactic protein 2) (SCYA8) gene of an individual, comprising determining for the two copies of the SCYA8 gene present in the individual the identity of the nucleotide pair at one or more polymorphic sites (PS) selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11 and PS12, wherein the one or more PS have the position and alternative alleles shown in SEQ ID NO:1.

6. The method of claim 5, wherein the determining step comprises:
   (a) isolating from the individual a nucleic acid mixture comprising both copies of the SCYA8 gene, or a fragment thereof, that are present in the individual;
   (b) amplifying from the nucleic acid mixture a target region containing one of the selected

polymorphic sites;

(c) hybridizing a primer extension oligonucleotide to one allele of the amplified target region, wherein the oligonucleotide is designed for genotyping the selected polymorphic site in the target region;

(d) performing a nucleic acid template-dependent, primer extension reaction on the hybridized oligonucleotide in the presence of at least one terminator of the reaction, wherein the terminator is complementary to one of the alternative nucleotides present at the selected polymorphic site; and

(e) detecting the presence and identity of the terminator in the extended oligonucleotide.

7. The method of claim 5, which comprises determining for the two copies of the SCYA8 gene present in the individual the identity of the nucleotide pair at each of PS1-PS12.

8. A method for haplotyping the small inducible cytokine subfamily A (Cys-Cys), member 8 (monocyte chemotactic protein 2) (SCYA8) gene of an individual which comprises determining, for one copy of the SCYA8 gene present in the individual, the identity of the nucleotide at two or more polymorphic sites (PS) selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11 and PS12, wherein the selected PS have the position and alternative alleles shown in SEQ ID NO:1.

9. The method of claim 8, wherein the determining step comprises:

(a) isolating from the individual a nucleic acid sample containing only one of the two copies of the SCYA8 gene, or a fragment thereof, that is present in the individual;

(b) amplifying from the nucleic acid sample a target region containing one of the selected polymorphic sites;

(c) hybridizing a primer extension oligonucleotide to one allele of the amplified target region, wherein the oligonucleotide is designed for haplotyping the selected polymorphic site in the target region;

(d) performing a nucleic acid template-dependent, primer extension reaction on the hybridized oligonucleotide in the presence of at least one terminator of the reaction, wherein the terminator is complementary to one of the alternative nucleotides present at the selected polymorphic site; and

(e) detecting the presence and identity of the terminator in the extended oligonucleotide.

10. A method for predicting a haplotype pair for the small inducible cytokine subfamily A (Cys-Cys), member 8 (monocyte chemotactic protein 2) (SCYA8) gene of an individual comprising:

(a) identifying a SCYA8 genotype for the individual, wherein the genotype comprises the nucleotide pair at two or more polymorphic sites (PS) selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11 and PS12, wherein the selected PS have the position and alternative alleles shown in SEQ ID NO:1;

(b) comparing the genotype to the haplotype pair data set forth in the table immediately

below; and

(c) determining which haplotype pair is consistent with the genotype of the individual and with the haplotype pair data

Haplotype Pair[a] (Part 1)								PS	PS
5/5	10/10	5/7	5/8	1/2	5/4	5/11	5/3	No.[b]	Position[c]
C/C	T/T	C/C	C/C	C/C	C/C	C/T	C/C	1	3499
G/G	G/G	G/G	G/G	A/A	G/G	G/G	G/G	2	3681
T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/C	3	3901
A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/C	4	3953
G/G	G/G	G/G	G/G	G/G	G/C	G/G	G/G	5	3961
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	6	3996
A/A	A/A	A/A	A/G	A/A	A/A	A/A	A/G	7	4019
C/C	C/C	C/G	C/C	C/C	C/C	C/C	C/C	8	4803
C/C	C/C	C/C	C/C	C/T	C/C	C/C	C/C	9	4848
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	10	4932
A/A	A/A	A/A	A/C	C/C	A/C	A/C	A/C	11	5381
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	12	5396

Haplotype Pair[a] (Part 2)				PS	PS
5/9	5/10	5/1	5/6	No.[b]	Position[c]
C/C	C/T	C/C	C/C	1	3499
G/G	G/G	G/A	G/G	2	3681
T/T	T/T	T/T	T/T	3	3901
A/A	A/A	A/A	A/A	4	3953
G/G	G/G	G/G	G/G	5	3961
C/T	C/C	C/C	C/C	6	3996
A/A	A/A	A/A	A/A	7	4019
C/C	C/C	C/C	C/C	8	4803
C/C	C/C	C/C	C/C	9	4848
C/A	C/C	C/C	C/C	10	4932
A/A	A/A	A/C	A/A	11	5381
G/G	G/G	G/G	G/T	12	5396

[a]Haplotype pairs are represented as 1st haplotype/2nd haplotype; with alleles of each haplotype shown 5′ to 3′ as 1st polymorphism/2nd polymorphism in each column;
[b]PS = polymorphic site;
[c]Position of PS in SEQ ID NO:1.

11. The method of claim 10, wherein the identified genotype of the individual comprises the nucleotide pair at each of PS1-PS12, which have the position and alternative alleles shown in SEQ ID NO:1.

12. A method for identifying an association between a trait and at least one haplotype or haplotype pair of the small inducible cytokine subfamily A (Cys-Cys), member 8 (monocyte chemotactic protein 2) (SCYA8) gene which comprises comparing the frequency of the haplotype or haplotype pair in a population exhibiting the trait with the frequency of the haplotype or haplotype pair in a reference population, wherein the haplotype is selected from haplotypes 1-11 shown in the table presented immediately below, wherein each of the haplotypes comprises a sequence of polymorphisms whose positions and identities are set forth in the table immediately

below:

Haplotype Number[a]											PS	PS
1	2	3	4	5	6	7	8	9	10	11	No.[b]	Position[c]
C	C	C	C	C	C	C	C	C	T	T	1	3499
A	A	G	G	G	G	G	G	G	G	G	2	3681
T	T	C	T	T	T	T	T	T	T	T	3	3901
A	A	C	A	A	A	A	A	A	A	A	4	3953
G	G	G	C	G	G	G	G	G	G	G	5	3961
C	C	C	C	C	C	C	C	T	C	C	6	3996
A	A	G	A	A	A	A	G	A	A	A	7	4019
C	C	C	C	C	C	G	C	C	C	C	8	4803
C	T	C	C	C	C	C	C	C	C	C	9	4848
C	C	C	C	C	C	C	C	A	C	C	10	4932
C	C	C	C	A	A	A	C	A	A	C	11	5381
G	G	G	G	G	T	G	G	G	G	G	12	5396

[a]Alleles for haplotypes are presented 5′ to 3′ in each column;
[b]PS = polymorphic site;
[c]Position of PS in SEQ ID NO:1;

and wherein the haplotype pair is selected from the haplotype pairs shown in the table immediately below, wherein each of the SCYA8 haplotype pairs consists of first and second haplotypes which comprise first and second sequences of polymorphisms whose positions and identities are set forth in the table immediately below:

Haplotype Pair[a]				(Part 1)				PS	PS
5/5	10/10	5/7	5/8	1/2	5/4	5/11	5/3	No.[b]	Position[c]
C/C	T/T	C/C	C/C	C/C	C/C	C/T	C/C	1	3499
G/G	G/G	G/G	G/G	A/A	G/G	G/G	G/G	2	3681
T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/C	3	3901
A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/C	4	3953
G/G	G/G	G/G	G/G	G/G	G/C	G/G	G/G	5	3961
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	6	3996
A/A	A/A	A/A	A/G	A/A	A/A	A/A	A/G	7	4019
C/C	C/C	C/G	C/C	C/C	C/C	C/C	C/C	8	4803
C/C	C/C	C/C	C/C	C/T	C/C	C/C	C/C	9	4848
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	10	4932
A/A	A/A	A/A	A/C	C/C	A/C	A/C	A/C	11	5381
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	12	5396

Haplotype Pair[a] (Part 2)				PS	PS
5/9	5/10	5/1	5/6	No.[b]	Position[c]
C/C	C/T	C/C	C/C	1	3499
G/G	G/G	G/A	G/G	2	3681
T/T	T/T	T/T	T/T	3	3901
A/A	A/A	A/A	A/A	4	3953
G/G	G/G	G/G	G/G	5	3961
C/T	C/C	C/C	C/C	6	3996
A/A	A/A	A/A	A/A	7	4019
C/C	C/C	C/C	C/C	8	4803
C/C	C/C	C/C	C/C	9	4848
C/A	C/C	C/C	C/C	10	4932
A/A	A/A	A/C	A/A	11	5381
G/G	G/G	G/G	G/T	12	5396

[a]Haplotype pairs are represented as 1st haplotype/2nd haplotype; with alleles of each haplotype shown 5′ to 3′ as 1st polymorphism/2nd polymorphism in each column;
[b]PS = polymorphic site;
[c]Position of PS in SEQ ID NO:1;

wherein a higher frequency of the haplotype or haplotype pair in the trait population than in the reference population indicates the trait is associated with the haplotype or haplotype pair.

13. The method of claim 12, wherein the trait is a clinical response to a drug targeting SCYA8.

14. An isolated oligonucleotide designed for detecting a polymorphism in the small inducible cytokine subfamily A (Cys-Cys), member 8 (monocyte chemotactic protein 2) (SCYA8) gene at a polymorphic site (PS) selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11 and PS12, wherein the selected PS have the position and alternative alleles shown in SEQ ID NO:1.

15. The isolated oligonucleotide of claim 14, which is an allele-specific oligonucleotide that specifically hybridizes to an allele of the SCYA8 gene at a region containing the polymorphic site.

16. The allele-specific oligonucleotide of claim 15, which comprises a nucleotide sequence selected from the group consisting of SEQ ID NOS:4-15, the complements of SEQ ID NOS:4-15, and SEQ ID NOS:16-39.

17. The isolated oligonucleotide of claim 14, which is a primer-extension oligonucleotide.

18. The primer-extension oligonucleotide of claim 17,which comprises a nucleotide sequence selected from the group consisting of SEQ ID NOS:40-63.

19. A kit for haplotyping or genotyping the small inducible cytokine subfamily A (Cys-Cys), member 8 (monocyte chemotactic protein 2) (SCYA8) gene of an individual, which comprises a set of oligonucleotides designed to haplotype or genotype each of polymorphic sites (PS) PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11 and PS12, wherein the selected PS have the position and alternative alleles shown in SEQ ID NO:1.

20. An isolated polynucleotide comprising a nucleotide sequence selected from the group consisting

of:

(a) a first nucleotide sequence which comprises a small inducible cytokine subfamily A (Cys-Cys), member 8 (monocyte chemotactic protein 2) (SCYA8) isogene, wherein the SCYA8 isogene is selected from the group consisting of isogenes 1- 4 and 6 - 11 shown in the table immediately below and wherein each of the isogenes comprises the regions of SEQ ID NO:1 shown in the table immediately below and wherein each of the isogenes 1- 4 and 6 - 11 is further defined by the corresponding sequence of polymorphisms whose positions and identities are set forth in the table immediately below; and

Isogene Number[a]										PS	PS	Regions
1	2	3	4	6	7	8	9	10	11	No.[b]	Position[c]	Examined[d]
C	C	C	C	C	C	C	C	T	T	1	3499	3321-4333
A	A	G	G	G	G	G	G	G	G	2	3681	3321-4333
T	T	C	T	T	T	T	T	T	T	3	3901	3321-4333
A	A	C	A	A	A	A	A	A	A	4	3953	3321-4333
G	G	G	C	G	G	G	G	G	G	5	3961	3321-4333
C	C	C	C	C	C	C	T	C	C	6	3996	3321-4333
A	A	G	A	A	A	G	A	A	A	7	4019	3321-4333
C	C	C	C	C	G	C	C	C	C	8	4803	4573-5115
C	T	C	C	C	C	C	C	C	C	9	4848	4573-5115
C	C	C	C	C	C	C	A	C	C	10	4932	4573-5115
C	C	C	C	A	A	C	A	A	C	11	5381	5201-5801
G	G	G	G	T	G	G	G	G	G	12	5396	5201-5801

[a]Alleles for isogenes are presented 5′ to 3′ in each column
[b]PS = polymorphic site;
[c]Position of PS in SEQ ID NO:1;
[d]Region examined represents the nucleotide positions defining the start and stop positions within the 1st SEQ ID NO of the sequenced region.

(b) a second nucleotide sequence which is complementary to the first nucleotide sequence.

21. The isolated polynucleotide of claim 20, which is a DNA molecule and comprises both the first and second nucleotide sequences and further comprises expression regulatory elements operably linked to the first nucleotide sequence.

22. A recombinant nonhuman organism transformed or transfected with the isolated polynucleotide of claim 20, wherein the organism expresses a SCYA8 protein that is encoded by the first nucleotide sequence.

23. The recombinant nonhuman organism of claim 22, which is a transgenic animal.

24. An isolated fragment of a small inducible cytokine subfamily A (Cys-Cys), member 8 (monocyte chemotactic protein 2) (SCYA8) isogene, wherein the fragment comprises at least 10 nucleotides in one of the regions of SEQ ID NO:1 shown in the table immediately below and wherein the fragment comprises one or more polymorphisms selected from the group consisting of thymine at PS1, adenine at PS2, cytosine at PS3, cytosine at PS4, cytosine at PS5, thymine at PS6, guanine at PS7, guanine at PS8, thymine at PS9, adenine at PS10, cytosine at PS11 and

thymine at PS12, wherein the selected polymorphism has the position set forth in the table immediately below:

Isogene Number[a]										PS	PS	Regions
1	2	3	4	6	7	8	9	10	11	No.[b]	Position[c]	Examined[d]
C	C	C	C	C	C	C	C	T	T	1	3499	3321-4333
A	A	G	G	G	G	G	G	G	G	2	3681	3321-4333
T	T	C	T	T	T	T	T	T	T	3	3901	3321-4333
A	A	C	A	A	A	A	A	A	A	4	3953	3321-4333
G	G	G	C	G	G	G	G	G	G	5	3961	3321-4333
C	C	C	C	C	C	C	T	C	C	6	3996	3321-4333
A	A	G	A	A	A	G	A	A	A	7	4019	3321-4333
C	C	C	C	C	G	C	C	C	C	8	4803	4573-5115
C	T	C	C	C	C	C	C	C	C	9	4848	4573-5115
C	C	C	C	C	C	C	A	C	C	10	4932	4573-5115
C	C	C	C	A	A	C	A	A	C	11	5381	5201-5801
G	G	G	G	T	G	G	G	G	G	12	5396	5201-5801

[a]Alleles for isogenes are presented 5′ to 3′ in each column
[b]PS = polymorphic site;
[c]Position of PS in SEQ ID NO:1;
[d]Region examined represents the nucleotide positions defining the start and stop positions within the 1st SEQ ID NO of the sequenced region.

.

25. An isolated polynucleotide comprising a SCYA8 coding sequence, wherein the coding sequence is selected from the group consisting of 1-4, 6, 8, 9 and 11 shown in the table immediately below, and wherein each of the coding sequences comprises SEQ ID NO:2, except at each of the polymorphic sites which have the positions in SEQ ID NO:2 and polymorphisms set forth in the table immediately below:

Coding Sequence Haplotype Number				PS	PS
1c,3c,4c,8c,11c	2c	6c	9c	No.[b]	Position[c]
C	T	C	C	9	117
C	C	C	A	10	201
C	C	A	A	11	235
G	G	T	G	12	250

[a]Alleles for the isogene coding sequence are presented 5′ to 3′ in each column; the numerical portion of the isogene coding sequence number represents the number of the parent full SCYA8 isogene;
[b]PS = polymorphic site;
[c]Position of PS in SEQ ID NO:2.

26. A recombinant nonhuman organism transformed or transfected with the isolated polynucleotide of claim 25, wherein the organism expresses a small inducible cytokine subfamily A (Cys-Cys), member 8 (monocyte chemotactic protein 2) (SCYA8) protein that is encoded by the polymorphic variant sequence.

27. The recombinant nonhuman organism of claim 26, which is a transgenic animal.

28. An isolated fragment of a SCYA8 cDNA, wherein the fragment comprises one or more

polymorphisms selected from the group consisting of thymine at a position corresponding to nucleotide 117, adenine at a position corresponding to nucleotide 201, cytosine at a position corresponding to nucleotide 235 and thymine at a position corresponding to nucleotide 250 in SEQ ID NO:2.

29 An isolated polypeptide comprising an amino acid sequence which is a polymorphic variant of a reference sequence for the small inducible cytokine subfamily A (Cys-Cys), member 8 (monocyte chemotactic protein 2) (SCYA8) protein, wherein the reference sequence comprises SEQ ID NO:3 for the regions encoded by SEQ ID NO:2, except the polymorphic variant comprises one or more variant amino acids selected from the group consisting of glutamine at a position corresponding to amino acid position 79 and phenylalanine at a position corresponding to amino acid position 84.

30. An isolated monoclonal antibody specific for and immunoreactive with the isolated polypeptide of claim 29.

31. A method for screening for drugs targeting the isolated polypeptide of claim 29 which comprises contacting the SCYA8 polymorphic variant with a candidate agent and assaying for binding activity.

32. An isolated fragment of the SCYA8 polypeptide, wherein the fragment comprises one or more variant amino acids selected from the group consisting of glutamine at a position corresponding to amino acid position 79 and phenylalanine at a position corresponding to amino acid position 84 in SEQ ID NO:3.

33 A computer system for storing and analyzing polymorphism data for the small inducible cytokine subfamily A (Cys-Cys), member 8 (monocyte chemotactic protein 2) gene, comprising:

(a) a central processing unit (CPU);

(b) a communication interface;

(c) a display device;

(d) an input device; and

(e) a database containing the polymorphism data;

wherein the polymorphism data comprises any one or more of the haplotypes set forth in the table immediately below:

Haplotype Number[a]											PS No.[b]	PS Position[c]
1	2	3	4	5	6	7	8	9	10	11		
C	C	C	C	C	C	C	C	C	T	T	1	3499
A	A	G	G	G	G	G	G	G	G	G	2	3681
T	T	C	T	T	T	T	T	T	T	T	3	3901
A	A	C	A	A	A	A	A	A	A	A	4	3953
G	G	G	C	G	G	G	G	G	G	G	5	3961
C	C	C	C	C	C	C	C	T	C	C	6	3996
A	A	G	A	A	A	A	G	A	A	A	7	4019
C	C	C	C	C	C	G	C	C	C	C	8	4803
C	T	C	C	C	C	C	C	C	C	C	9	4848
C	C	C	C	C	C	C	C	A	C	C	10	4932
C	C	C	C	A	A	A	C	A	A	C	11	5381
G	G	G	G	G	T	G	G	G	G	G	12	5396

[a]Alleles for haplotypes are presented 5′ to 3′ in each column;
[b]PS = polymorphic site;
[c]Position of PS in SEQ ID NO:1;

the haplotype pairs set forth in the table immediately below:

Haplotype Pair[a] (Part 1)								PS No.[b]	PS Position[c]
5/5	10/10	5/7	5/8	1/2	5/4	5/11	5/3		
C/C	T/T	C/C	C/C	C/C	C/C	C/T	C/C	1	3499
G/G	G/G	G/G	G/G	A/A	G/G	G/G	G/G	2	3681
T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/C	3	3901
A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/C	4	3953
G/G	G/G	G/G	G/G	G/G	G/C	G/G	G/G	5	3961
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	6	3996
A/A	A/A	A/A	A/G	A/A	A/A	A/A	A/G	7	4019
C/C	C/C	C/G	C/C	C/C	C/C	C/C	C/C	8	4803
C/C	C/C	C/C	C/C	C/T	C/C	C/C	C/C	9	4848
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	10	4932
A/A	A/A	A/A	A/C	C/C	A/C	A/C	A/C	11	5381
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	12	5396

Haplotype Pair[a] (Part 2)				PS No.[b]	PS Position[c]
5/9	5/10	5/1	5/6		
C/C	C/T	C/C	C/C	1	3499
G/G	G/G	G/A	G/G	2	3681
T/T	T/T	T/T	T/T	3	3901
A/A	A/A	A/A	A/A	4	3953
G/G	G/G	G/G	G/G	5	3961
C/T	C/C	C/C	C/C	6	3996
A/A	A/A	A/A	A/A	7	4019
C/C	C/C	C/C	C/C	8	4803
C/C	C/C	C/C	C/C	9	4848
C/A	C/C	C/C	C/C	10	4932
A/A	A/A	A/C	A/A	11	5381
G/G	G/G	G/G	G/T	12	5396

[a]Haplotype pairs are represented as 1st Haplotype/2nd Haplotype; with alleles of each haplotype shown 5′ to 3′ as 1st polymorphism/2nd polymorphism in each column;
[c]Position of PS in SEQ ID NO:1;

and the frequency data in Tables 6 and 7.

34. A genome anthology for the small inducible cytokine subfamily A (Cys-Cys), member 8 (monocyte chemotactic protein 2) (SCYA8) gene which comprises two or more SCYA8 isogenes selected from the group consisting of isogenes 1-11 shown in the table immediately below, and wherein each of the isogenes comprises the regions of SEQ ID NO:1 shown in the table immediately below and wherein each of the isogenes 1-11 is further defined by the corresponding sequence of polymorphisms whose positions and identities are set forth in the table immediately below:

Isogene Number[a]											PS	PS	Regions
1	2	3	4	5	6	7	8	9	10	11	No.[b]	Position[c]	Examined[d]
C	C	C	C	C	C	C	C	C	T	T	1	3499	3321-4333
A	A	G	G	G	G	G	G	G	G	G	2	3681	3321-4333
T	T	C	T	T	T	T	T	T	T	T	3	3901	3321-4333
A	A	C	A	A	A	A	A	A	A	A	4	3953	3321-4333
G	G	G	C	G	G	G	G	G	G	G	5	3961	3321-4333
C	C	C	C	C	C	C	C	T	C	C	6	3996	3321-4333
A	A	G	A	A	A	A	G	A	A	A	7	4019	3321-4333
C	C	C	C	C	C	G	C	C	C	C	8	4803	4573-5115
C	T	C	C	C	C	C	C	C	C	C	9	4848	4573-5115
C	C	C	C	C	C	C	C	A	C	C	10	4932	4573-5115
C	C	C	C	A	A	A	C	A	A	C	11	5381	5201-5801
G	G	G	G	G	T	G	G	G	G	G	12	5396	5201-5801

[a]Alleles for isogenes are presented 5′ to 3′ in each column
[b]PS = polymorphic site;
[c]Position of PS in SEQ ID NO:1;
[d]Region examined represents the nucleotide positions defining the start and stop positions within the 1st SEQ ID NO of the sequenced region.

1/6

POLYMORPHISMS IN THE SCYA8 GENE

GGCCACTTTG GGCAGGTTAT TGGTGAAATC AAGACTAGAC TCAAGTTCCT
TGACACTTAA TTCCCTTCAT TTTTACTAGC CTTTTTTGAT TCCCAATAAT 100
TATAATTTTC ATGACTATTT CTGTTGTCTG TGCATGGATG CACTTTTCGT
TTGAAGTCAG CTCAGAGGAA AATTAACAAA GTCCTCAGGG TATTACATAA 200
GACAATGGCT TTATTGCTGC TTGGAAAATA CACACAAACT CCACCCCTAA
ACCAATACCC CCACCCCATG CACACACTCA GATTGATGCT ACACAGTACA 300
CCTGGCAGAG AACCATCCCC TAAAATGATT AAAAAGCTGC CATCCTGAAA
CAACTCAATT TCCTTTCTGA TTAAAGGAAA TCCAAATGAC ACACATGTTC 400
CATATACTGC AAATTCTTTA CTGTTAAAAG TTCACATAGT AATTGTAATA
AATGAAAAAT GCATGAGGTA AATGTTTAAA AATCTCTTCT GAATCAAAAC 500
AGTAAAAACC CACCTGATTT AAGAACAGGA GCATATTAAC CACAAGAAAA
AAATAAAGCA ACTAACATAC ATATTCAAAA GCAAATAAAG TAAAATGTAA 600
GCAAAATAAT GGTGAGCCTT TTGCTTCGAT TTACTTTATT TCGAAGTATA
ATTCAATATA CTGTCCAGTA AAGTGCAGGC ATTTTAAATG TAGAGTTTCA 700
TGAATTTTAC ATGTGTAGTT AACCATTTAA CTACCTCTCT GATCAAAATC
AGAGAGGGAA GCTTCCAAAA GTTTCCCTCA TGATGTCCAT ACCTGCAAAG 800
GAACCACTGT TCTAACCTCT ATCACTGTAC GTTTCTTTTG TCTTTTAACA
AGAAAAATAT CTTCTTCTTA GCAACTATAA AATGTTTTTA AGAATAATAC 900
ACTCACATTT AAGTGAAGGG GCAAATTTTC CAGCAATGTG TTCATTTGAT
TAAAAGTCTT AATAATATCC ACATATTTTG ATTTATTAAT TCTTTGGCTA 1000
GAAAATTAAA CGAAGACATT AATAATAATG TTATATCAGC TTTTCCACAA
GTATTGTGTA AAACCATAAA AATGGAAATG GTATGATTCC CAACTATGGA 1100
GAAGTAAATT CTTAAGTTAA TTATGGTGCA CTGATATATA TGAACTCCTA
TAGAGCTTTT AAAGTTCAAG ATTATTGGGG AATATTTACT GTCTTGTGAG 1200
AAATCTAAAC ACACTAATTG AAAAGGCAGT ATTCAATCTT GGATTTGGTG
GTGAAATTTT AGATATACAC CCTATAGTTC AACAATACAT TGCCAATTAC 1300
TAACCGATGT TATTTCTATA AAACAATGAG AATTCCTGAG GAACAAATTT
TGTAATTGCC CCCTCTCCTC CCTTCCTAAT TCATTCTGAG TCCAGAATGA 1400
CCCTAACATC AAAACCAGAT AAAGACATTA CAACAAAAGA AAACTATAAA
CCAACAGAAA TCTATAAACC AAACTCATGA ACAAAATTCT CAACAAAATA 1500
TCAGCAAATA AAATCCGACA ACATATAAAA AGAATTATAA GCTACAACCA
AGTGGAATTT ATTCCAGGTA TGCAAGGCTG ATCCAGCGTT CAAAAATTGG 1600
TTGATATAAC CCATCATATC AACTTTCTAA AGAAGGAAAA TCTTATTATA
ATATCAATAG ATGGAAAATT TTTTTTGACA CAATTCAACA GCCATTCATG 1700
ATTTAAAAAA AAAAACTCTC AGCAAACTAG GAATAGAGGG ATACTTCCTC
AGCTTTATAG AGTGCATTAA AAAAGACCTA CAGCAAATAT CATACTTAAT 1800
GGTGAGAAAC TAGATATTTT TCTCCTAAGA TTGAGTATAA GGCAAGGATG
TTTCTTCTCA CCACTCCAAT TCAACATAAT ACTGGAAGGC CTAGCCAACA 1900
CAACAGCACA AGACAAGAAA ATAAAAATAT ATACAGATTA GAAAGAATAA
ATAAAACTGA ATTTATTTGC AGAAGACATG ATTGTCTGTG TAGAAAATTA 2000
CAGAGAATCA GCAACAAAAA AAAGTCATGC CACTAATAAG CAATTTTAGC
TACAGGATAC AAGGCTAATA GATAAAATTC AATAACTTCC TGTATATCAG 2100
CAATGAGCAA CTAGAATTTG AAATTAATAC ACAATACTAT TTACATTAAC
ACCAATAAAA TGAAGTGTTT AGAAATTAGT CTAAGATGTA CAATATATAT 2200
GAGGGAAACT ACAAACTTCT GATGAAAGAT TAAAGACGAT CTAAATAAAT
GAAGAGATAT TCCACATGTA TGGATAGGTG GAGTTAATGT TGTCAGCATG 2300
TCAGTTCTTC CCAACTTGAT CTGTACCTTC AACACAATCC CAGTCCAAAT
CCCAGCAGGT TCTTTTTGGA TGTGATAGAC TGATTTTAAT GTTTATATGG 2400
AAAGGTTAAA GATCTAGAAT AGCTAACACA GTACTGAAAA AGAATAGTCA
GAGGACTAAC ATTACCCAAC TTCAAGATTT ACTATTAGGA AACAGTAATC 2500
AAGACGGTGT GGTATTGGTG AAAGAATAGA CAAATGGATC AATGGAACAG

FIGURE 1A

```
AATAGAGAGC CCTCAACAGA TGAATGGATA CATAAAATGT ACTATATACA     2600
TACAAGGGAA TATTACTCAG CCTTAAGGAT ATGTAGCATA TCATATGCTG
CCACATAGAT GAACCCTGAG ATCCTTACGC TAAATGACAT AAGCCAGTCA     2700
CAAAAGAAAA ATACTATATG ATTTCACTCA CATGAGGTAG TTAGAGCAGT
CAAAATCAGA AACAGTAAGT AGAAATGGTG GTTGCCAGGG GCTTGGGGGA     2800
GGGGAAACTG GAGATCAATA GGTATAGAAT TTCTGTTTAC AGATGAAAAG
AGTTATTGAG ACTGTTGCAC AACAGTGTGA ATGTTCTTAA TGCCACTAAA     2900
CTATGCACTT AAAAATGGTT AAGATGGTAT GTGCATTTTT ACCACAATAA
AAATAATAAA TGAATAAAGA ACAAAGGAAG TAATTAGCTA TCAAACCACA     3000
ACAAGGCACG GAGAAACCTT AAATGCATAT TGCTTAGTGA AAGAAGCCAG
TCTGAAAAGG CTGCCTACTA TATGATTCCA ACTATGCATA TGATACTTGG     3100
GAAAAGACAA GCTCACAGAG ATAGTAAAAA GATCAGTGGT AGCCAGAGAT
TTGTGGAGAG GAAAAAGGAT GAATAGAAGA ATACAGTGAA TTCTTAGAGC     3200
AGGAAAACTT CTATATGATA TTGTAATGAT GGATAGATCA GGCCTGCAGG
GGGAGTAGAT GAGGTCAGTC TTTTGCAAAC CTGGTTCCTT CTTGCCACGC     3300
CCCGCCCACA GCCCCACTGT GTGTGAACCA AGAGAGTTGG AAGGCCTTCT
GCAGACGGAC CCATTCCACC TGAATCTCCA TGACATATTG GACTTCTCCA     3400
TATAATTTCA TTTGGAAAGA AACTTGAGCA TCTAAATAAA ATGTTTGGAA
ACCACCATGC TACAAGACCA GCATGTGCTG TCTAGTTCTT ACCCTCTGCA     3500
                                                 T
ACAGTAAGAT TCTGGGGCAT TAAGACTTAG TTCCAGGATT CTGTCATTCT
GCCAACGTTC TGTGGCTGGG GTTCTAAAGG AGCTTGCCTG GCTTAGAACT     3600
GCAAGTGACT CTAGTGTGAT GGAGAGCACC AGCAAAGCCT TAGGGCCCAT
CCCTGGCCTC CTGTTACCCA CAGAGGGGTA GGCCCTTGGC TCTCTTCCAC     3700
                                 A
TATGACGTCA GCTTCCATTC TTCCTTTCTT ATAGACAATT TTCCATTTCA
AGGAAATCAG AGCCCTTAAT AGTTCAGTGA GGTCACTTTG CTGAGCACAA     3800
TCCCATACCC TTCAGCCTCT GCTCCACAGA GCCTAAGCAA AAGATAGAAA
CTCACAACTT CCTTGTTTTG TTATCTGGAA ATTATCCCAG GATCTGGTGC     3900
TTACTCAGCA TATTCAAGGA AGGTCTTACT TCATTCTTCC TTGATTGTGA
C
CCATGCCCAG GCTCTCTGCT CCCTATAAAA GGCAGGCAGA GCCACCGAGG     4000
  C         C                                 T
AGCAGAGAGG TTGAGAACAA CCCAGAAACC TTCACCTCTC ATGCTGAAGC
                   G
     [exon 1: 4041..
TCACACCCTT GCCCTCCAAG ATGAAGGTTT CTGCAGCGCT TCTGTGCCTG     4100
CTGCTCATGG CAGCCACTTT CAGCCCTCAG GGACTTGCTC AGCCAGGTAA
                ..4146]
GACCTCTCCC TTTTTAAGGG GAGACCAAAA GAGGAATTAA GAAGAGCCAT     4200
TATGTCACAG CTCATTAGGA ACAAAAACCAG AACTAAAGGC TCAGGTCACT
GAGGCTGGTT CCCTTGATCT TTCCTGACCC CAGTTTTGGG AGGAGACAGT     4300
GGAGCCGCTA CAGCAACAAC CCTCCCATTG TTTGGGGAAA TAATCCAGAA
CGAAGAACTG TTTCTCACTG TGGGTGTAAA GGACATTTCA GGCCGTAGTG     4400
GAGAGGGAGA AACTATTGCC TGAAGCTTCA AATTTTGGTT ATGGTTCAGT
GTACCTTCCA GAACAGTGGC TGTGTAAAGA GGATGAGGAC CCAGAGGAAT     4500
CTCAGCGTAT GGCATAGGCT AACTCTAAAG CCCATGAGGA TGAAAGACTG
GGAAGCAAGG TATTGGAACT TATGTTCCCA GTGTCAGAAG TTTTGGGTTA     4600
GTAGACAAGG ACTAGCTTGT TACTCAAAAT GTTTCCAAAC CCAGTCAACA
ATGACGGGCC GCAGAGTTCA ATAGAGGAAA GAGACTCACA GGCAACATTT     4700
TATCTCTGGG ATCTGGACTA AGACACTGAA CTTGGGATGG TGACTTCTTG
GTCTTCTCCT TCCTTCTCTT CTTTTCCTTA CAAATGCACA CTTACGGTGG     4800
```

FIGURE 1B

```
GTCCTAAATG TCTCATTCTT TGCAAAATTT CTTTCAGATT CAGTTTCCAT
  G                                                 T
     [exon 2: 4838..
TCCAATCACC TGCTGCTTTA ACGTGATCAA TAGGAAAATT CCTATCCAGA        4900
GGCTGGAGAG CTACACAAGA ATCACCAACA TCCAATGTCC CAAGGAAGCT
                                 A
GTGATGTGAG TGGACAGTGC CTGGCACCCC CATTCAAAAG TTCTGATGGA        5000
                ..4955]
CAACATAGAG AAGTCAAGAT TCATGTCCAT ATGAGTCGGA TGCATATAAC
TTCTATCCAA AGGGCCCCTC TACCCCATAG AGAAACTCAG TCCATGAGAA        5100
GGAGTCCATA ACTGCTCTAG GATTCCCTTC TAGGGGCTTG GTGAAACTAA
CCCAATATCT GTAGCCAGGA CCCTGGAGGG TTTCACCTGG ACAGCAAGAG        5200
CAGAGCTTCC TTCTGGAGCT TCTTCCTCCC ACTCTTCCCC TCCCTCCTCT
CCCGGGTCCG GGTCCTTCAC CTAAGGACCA AGGGCTGATC AGTTCTAGGG        5300
ACCAATGGCC CACAGTCCTG TGCAGGATCT TCAAAGTCTT CCATCTAATT
GTGCCCTCTC TCCCCCACAG CTTCAAGACC AAACGGGGCA AGGAGGTCTG        5400
                                 C              T
     [exon 3: 5371..
TGCTGACCCC AAGGAGAGAT GGGTCAGGGA TTCCATGAAG CATCTGGACC
AAATATTTCA AAATCTGAAG CCATGAGCCT TCATACATGG ACTGAGAGTC        5500
                ..5476]
AGAGCTTGAA GAAAAGCTTA TTTATTTTCC CCAACCTCCC CCAGGTGCAG
TGTGACATTA TTTTATTATA ACATCCACAA AGAGATTATT TTTAAATAAT        5600
TTAAAGCATA ATATTTCTTA AAAAGTATTT AATTATATTT AAGTTGTTGA
TGTTTTAACT CTATCTGTCA TACATCCTAG TGAATGTAAA ATGCAAAATC        5700
CTGGTGATGT GTTTTTTGTT TTTGTTTTCC TGTGAGCTCA ACTAAGTTCA
CGGCAAAATG TCATTGTTCT CCCTCCTACC TGTCTGTAGT GTTGTGGGGT        5800
CCTCCCATGG ATCATCAAGG TGAAACACTT TGGTATTCTT TGGCAATCAG
TGCTCCTGTA AGTCAAATGT GTGCTTTGTA CTGCTGTTGT TGAAATTGAT        5900
GTTACTGTAT ATAACTATGG AATTTTGAAA AAAAATTTCA AAAAGAAAAA
AATATATATA ATTTAAAACT ACTTAGTCTT ATTCTTCTTG GGGTAACATT        6000
TAGCTGGGAG TGAGTTTTGG GCATCATGGG TGACAGTTTG GGCATGGACG
GGCCATTTTT CAAGAATGTC TTCTGGCTAC GCTGGACTCA ACCAAGGTTC        6100
TCAGAGAACT TGGTGGGACC AGGCCAGGAT GTTCCAGCTC TCTGACTCTA
GTCCCTAACT TCAGCAGCCC TGATTCGCTA GCCTCTCTTG TTTCTCTTGT        6200
TTATATATTA TCCAGCCTAA GGTATTTTGT TATAGCTGCC CAAAAAGACT
AAGATAATCT CCATCACTCT ACCCCCAACC CCAATCCCAA GAACTTGCAA        6300
GCATCCATTT AAAGGCGTGG AACCTCTTCT TTTTGACAGC CTTTTAAGGT
CAAGATTCCC CTGTACTTAG TGAGCTTAGC TGAATCTTCT TACAAACATG        6400
TGACCCGCCA TATTGAGCCA TACATACCAG AGCTTATTAT TTTTCCAGCT
TATTGGGAAA ACACGTCTAA GGCAAACAAA TTTATTGTAC TGTTGAACCA        6500
ACTCTGCAGG GCCTACTGTT AATGTTCATG TAAAGGTGCC ACAATTCCAG
TATATTTGGA AGTATAGAGA GGAAAGAGAT TTGGGTTCTG CTCCTAAGAG        6600
TAACCTTAAC TCTCTTGGTC ACTTAGGGTG ACATAATTTT TCCCTTCAGC
ATTAAGATGC CCCATCCAAA AATCAGGATC ATACTGTGTC ACAGGCAGGA        6700
AGTGAGAGGA TCTAGCAGAT GACTACACAA GTCCCTCCAG CTCTAGAGCC
ATGTCTCTTA GAAACAAGGT CATGTTTGTT GTCTGGGGTA GAGTGGGGAC        6800
ATAAAGTTCC TGGGATTAAA GCCATAAAGG CAGCCAGGGC CTAGCTTATT
GGAAAAACAC ATCTAAGTCA GATACCCCTT ATTTGGAGAC AAGCTTTCCA        6900
TTAAAGTTCA AAGTAGTACA CCCACAGTTC CAAGAGCAAT AAAGCAGCTA
AGTGTCAGAT GGACTAGAGG GGCACCTCCC AGGCAGGGTC CACAAAAAAG        7000
CTCCATCAAA CTCTAAGCTC TTTCTTCAAT AGCTTCAAGG TCAATTGGGC
ACAGATATGT GGGTCCTGGG CTTCATGAAA TCCTGGACCA CAAATCCCAA        7100
```

FIGURE 1C

```
ATTCTAAAGG CAAGAACACT CACACAAAAC TGACACCAGG CCACAGCAAA
AAGACACAAA TTTGTTCTTC TCAAATTTTA CCCAGCTTCA CTTTGGGATT      7200
ATGTCATTAT ATTTCAAATG ACAATGAGCA TAATACACCT TAAGTAACAT
GTCATCTTAT TTAAGTGGTT GATGTTTTAT GTGCCCTGAC TTGAATATAT      7300
GTGTTTTAAA AATCCACATC CTTCCATAAA CTGGTTCAGA TCTGTAAGCC
CTGCTTTGCT GTCTTGTAAA CTGGGGGAAA GACCTGGAAA GGTCACTCAA      7400
AGCTGACATA CAACCTCAGG CTCACCTCAG TTTCACCCTC AGCTAAGGCT
GACCATTTGT GGCAGTCCAT GACTCACTTG ATTCACTCAA TATCTGCTTT      7500
CTGCCTCTTC TCACTTCTCA CTTGCTGCCC CCAGGCTTCT CCTGGGTGTC
ATCTTGGGTC AGGAGTTGAC TCCACTGCAC ACTTGTCCAA GTTTCAGAAG      7600
TGTGGAAGAG TGAACATTTT GGGTGGTAAC CGTCCACAGT GAAGGGTGAA
GGCAGATGTT CAAATGCTTC TGCCATCCCT TTTACCAGTT CTGCTCTGAT      7700
CTTGGTTATT TCTTTTCTTC TGCTGGGGTT GGATTTGGTT TGTTCTTGTT
TCTCTAGTTC CTTAAGGTGT GACATTAGAT TGTCTATTAA TGCTCTTTCA      7800
GACTTTTTCT TTTAAATATA GGCATTTAAT GCTATGAGCT TTCCTCTTAG
TACCGTTTTT GCTGTATCCC AAAGGTTTTG ATAGGTTGTG TCATTATCAT      7900
TGTTCCGTTC AAAGATTTAA AAAATTTTCA TCTTGATTTC ATTGTTGACC
CAACAATCAT TCAGGAATAC ATTATTTAAT TTCCATATAT TTGCATGGTT      8000
TTGAGGTTTC CTTTTGGAGT TGATTTCCAA TTTTATTCCA CTTTGGTCTG
AGAGAGTACT TGATATAATT TCAATTTTCT TAAATTTGTT GAGACTTGTT      8100
TTGTGGCCTA TCATAATGTC TATCTTGGAG AATGTTCCAT GTGCTGATGA
ATAGAATGTA TACTCTGCAG TTGTTGGGTA GAATGGTCTG TAAATATCCC      8200
TTAAGTCCAT TTGCTCTAGG GTATAGTTTA AGTCCATTGT TTCTTTGTTG
ACTTTCTGTC TTGATGACCT GTCTAGTGCT GCCATTGGAG TATTGAAGTC      8300
CCCCACTATT ATTGTGTTGC TATCTCATTT CTTAGGTCTA GTAGTAATTG
TTTTATAGAT TTGGGAGCTC CAGTGTTAGG TGCATATATA TTTAGGACTG      8400
ATATTTTCCT GTTGGACTAG TCCTTTTATC ATTATATAAT GTCTCTTTTT
GTCTTTTTTA ATTGCTGTTC CTTCAAAATT TGTTTTGTCT GATATAAAAA      8500
TAGCTACTCC TGCTCATTTT TGGTGTCCAT TTGTATGGAA TATCTTCTTC
CACCCCTCTA TCTTAAGCTC ATGTGAGTCC TAATGTGTCA GGTGAGTCCC      8600
CTGAAAACAG CAGATACTTG GTGAATTCTT ATCCATTCTG CCATTCTGTA
TTTTTGAAGT GGAGCATTTA GGCTATTTAC ATTCAATGTT AGTATTGAGA      8700
TGTGAGGTAC TACTCTATTC ATAATACCAT TTGTTGCCTG AATACCTTTT
TTTCATTGTG TTATTGTTTT CTAAGTCCTG TGAGATTTAT GCTTTCAGGA      8800
GATTCTAGTT TGGTGTATTT TGTGGATTTG TTTCAAGATT TAGAGCTCCT
TTTAGCAGTT CTTGTAGTGC CGGCTCGGCG GTGGCAAATT CTCTCTGCAT      8900
TTGTTTGTCT GGAAAAGACT ATCTTGCCTT CATTTATGAA GCTTAGTTTT
GCTGGATACA AAATTCTTGG CTGATAATTG TTTTGTTTAA GGGGGCTAAA      9000
TATAGAACCC CAATTCCTTC TAGCTTGTAG GGTTCCTGCT AAGAAATCTG
CTGTTAATCT GATAGGTTTT CTTTTATAGG TTACCTGGTA CTTTTGCCTC      9100
ACAGCTCTTA AGAGTCTTTC CTTCCTCTTG ACTTTAGATA ATCTGATGAC
TACTTGCCAA GGTGATGGTC TTTTTACAAT GAATTTCCCA GGTGTTCTTT      9200
GAGCTTCTTT TATTTGGATG TCTGAATCTC TAGCAAGTCT GAGGAAGTTT
TCCTCAAATA TGTTTTCTAA ACATATTTGG GGAACATATT TCTAAACATT      9300
CCCTCAAATA TGTTTTCTAA ACTTTTAGAT TTCTCTTCTT CCTTGGGAAC
ACCAATAATT CTTAGATTTG GTCGTTTAAC ATAATCTCAA GCTTCTTAGA      9400
GGCTATATTC ATTTTTTAAA TTCTTTTTTC TTCACCTTTG TTGGATTGGT
TTAATTCAAA AGCCTTGTCT TCAAG                                 9475
```

FIGURE 1D

POLYMORPHISMS IN THE CODING SEQUENCE OF SCYA8

```
ATGCTGAAGC TCACACCCTT GCCCTCCAAG ATGAAGGTTT CTGCAGCGCT
TCTGTGCCTG CTGCTCATGG CAGCCACTTT CAGCCCTCAG GGACTTGCTC      100
AGCCAGATTC AGTTTCCATT CCAATCACCT GCTGCTTTAA CGTGATCAAT
                 T
AGGAAAATTC CTATCCAGAG GCTGGAGAGC TACACAAGAA TCACCAACAT      200
CCAATGTCCC AAGGAAGCTG TGATCTTCAA GACCAAACGG GGCAAGGAGG
A                                    C               T
TCTGTGCTGA CCCCAAGGAG AGATGGGTCA GGGATTCCAT GAAGCATCTG      300
GACCAAATAT TTCAAAATCT GAAGCCATGA                            330
```

FIGURE 2

ISOFORMS OF THE SCYA8 PROTEIN

```
MLKLTPLPSK MKVSAALLCL LLMAATFSPQ GLAQPDSVSI PITCCFNVIN
RKIPIQRLES YTRITNIQCP KEAVIFKTKR GKEVCADPKE RWVRDSMKHL         100
                              Q     F
DQIFQNLKP                                                      109
```

FIGURE 3

SEQUENCE LISTING

<110> Genaissance Pharmaceuticals
Anastasio, Alison E.
Chew, Anne
Han, Jin-Hua
Lee, Helen H.

<120> Haplotypes of the SCYA8 Gene

<130> MWH-1304PCT

<140> TBA
<141> 2001-09-17

<150> 60/232,755
<151> 2000-09-15

<160> 64

<170> PatentIn Ver. 2.1

<210> 1
<211> 9475
<212> DNA
<213> Homo sapien

<220>
<221> allele
<222> (3499)
<223> PS1: polymorphic base C or T

<220>
<221> allele
<222> (3681)
<223> PS2: polymorphic base G or A

<220>
<221> allele
<222> (3901)
<223> PS3: polymorphic base T or C

<220>
<221> allele
<222> (3953)
<223> PS4: polymorphic base A or C

<220>

<221> allele
<222> (3961)
<223> PS5: polymorphic base G or C

<220>
<221> allele
<222> (3996)
<223> PS6: polymorphic base C or T

<220>
<221> allele
<222> (4019)
<223> PS7: polymorphic base A or G

<220>
<221> allele
<222> (4803)
<223> PS8: polymorphic base C or G

<220>
<221> allele
<222> (4848)
<223> PS9: polymorphic base C or T

<220>
<221> allele
<222> (4932)
<223> PS10: polymorphic base C or A

<220>
<221> allele
<222> (5381)
<223> PS11: polymorphic base A or C

<220>
<221> allele
<222> (5396)
<223> PS12: polymorphic base G or T

<400> 1
ggccactttg ggcaggttat tggtgaaatc aagactagac tcaagttcct tgacacttaa 60
ttcccttcat ttttactagc ctttttgat tcccaataat tataattttc atgactattt 120
ctgttgtctg tgcatggatg cacttttcgt ttgaagtcag ctcagaggaa aattaacaaa 180
gtcctcaggg tattacataa gacaatggct ttattgctgc ttggaaaata cacacaaact 240
ccacccctaa accaataccc ccaccccatg cacacactca gattgatgct acacagtaca 300
cctggcagag aaccatcccc taaaatgatt aaaaagctgc catcctgaaa caactcaatt 360
tcctttctga ttaaaggaaa tccaaatgac acacatgttc catatactgc aaattcttta 420
ctgttaaaag ttcacatagt aattgtaata aatgaaaaat gcatgaggta aatgtttaaa 480

aatctcttct gaatcaaaac agtaaaaacc cacctgattt aagaacagga gcatattaac 540
cacaagaaaa aaataaagca actaacatac atattcaaaa gcaaataaag taaaatgtaa 600
gcaaaataat ggtgagcctt ttgcttcgat ttactttatt tcgaagtata attcaatata 660
ctgtccagta aagtgcaggc attttaaatg tagagtttca tgaattttac atgtgtagtt 720
aaccatttaa ctacctctct gatcaaaatc agagagggaa gcttccaaaa gtttccctca 780
tgatgtccat acctgcaaag gaaccactgt tctaacctct atcactgtac gtttcttttg 840
tcttttaaca agaaaaatat cttcttctta gcaactataa aatgttttta agaataatac 900
actcacattt aagtgaaggg gcaaattttc cagcaatgtg ttcatttgat taaaagtctt 960
aataatatcc acatattttg atttattaat tctttggcta gaaaattaaa cgaagacatt 1020
aataataatg ttatatcagc ttttccacaa gtattgtgta aaaccataaa aatggaaatg 1080
gtatgattcc caactatgga gaagtaaatt cttaagttaa ttatggtgca ctgatatata 1140
tgaactccta tagagctttt aaagttcaag attattgggg aatatttact gtcttgtgag 1200
aaatctaaac acactaattg aaaaggcagt attcaatctt ggatttggtg gtgaaatttt 1260
agatatacac cctatagttc aacaatacat tgccaattac taaccgatgt tatttctata 1320
aaacaatgag aattcctgag gaacaaattt tgtaattgcc ccctctcctc ccttcctaat 1380
tcattctgag tccagaatga ccctaacatc aaaaccagat aaagacatta caacaaaaga 1440
aaactataaa ccaacagaaa tctataaacc aaactcatga acaaaattct caacaaaata 1500
tcagcaaata aaatccgaca acatataaaa agaattataa gctacaacca agtggaattt 1560
attccaggta tgcaaggctg atccagcgtt caaaaattgg ttgatataac ccatcatatc 1620
aactttctaa agaaggaaaa tcttattata atatcaatag atggaaaatt tttttgaca 1680
caattcaaca gccattcatg atttaaaaaa aaaaactctc agcaaactag gaatagaggg 1740
atacttcctc agctttatag agtgcattaa aaaagaccta cagcaaatat catacttaat 1800
ggtgagaaac tagatatttt tctcctaaga ttgagtataa ggcaaggatg tttcttctca 1860
ccactccaat tcaacataat actggaaggc ctagccaaca caacagcaca agacaagaaa 1920
ataaaaatat atacagatta gaaagaataa ataaaactga atttatttgc agaagacatg 1980
attgtctgtg tagaaaatta cagagaatca gcaacaaaaa aaagtcatgc cactaataag 2040
caattttagc tacaggatac aaggctaata gataaaattc aataacttcc tgtatatcag 2100
caatgagcaa ctagaatttg aaattaatac acaatactat ttacattaac accaataaaa 2160
tgaagtgttt agaaattagt ctaagatgta caatatatat gagggaaact acaaacttct 2220
gatgaaagat taaagacgat ctaaataaat gaagagatat tccacatgta tggataggtg 2280
gagttaatgt tgtcagcatg tcagttcttc ccaacttgat ctgtaccttc aacacaatcc 2340
cagtccaaat cccagcaggt tctttttgga tgtgatagac tgattttaat gtttatatgg 2400
aaaggttaaa gatctagaat agctaacaca gtactgaaaa agaatagtca gaggactaac 2460
attacccaac ttcaagattt actattagga aacagtaatc aagacggtgt ggtattggtg 2520
aaagaataga caaatggatc aatggaacag aatagagagc cctcaacaga tgaatggata 2580
cataaaatgt actatataca tacaagggaa tattactcag ccttaaggat atgtagcata 2640
tcatatgctg ccacatagat gaaccctgag atccttacgc taaatgacat aagccagtca 2700
caaaagaaaa atactatatg atttcactca catgaggtag ttagagcagt caaaatcaga 2760
aacagtaagt agaaatggtg gttgccaggg gcttgggga ggggaaactg gagatcaata 2820
ggtatagaat ttctgtttac agatgaaaag agttattgag actgttgcac aacagtgtga 2880
atgttcttaa tgccactaaa ctatgcactt aaaaatggtt aagatggtat gtgcattttt 2940
accacaataa aaataataaa tgaataaaga acaaaggaag taattagcta tcaaaccaca 3000
acaaggcacg gagaaacctt aaatgcatat tgcttagtga aagaagccag tctgaaaagg 3060
ctgcctacta tatgattcca actatgcata tgatacttgg gaaaagacaa gctcacagag 3120
atagtaaaaa gatcagtggt agccagagat ttgtggagag gaaaaaggat gaatagaaga 3180
atacagtgaa ttcttagagc aggaaaactt ctatatgata ttgtaatgat ggatagatca 3240
ggcctgcagg gggagtagat gaggtcagtc ttttgcaaac ctggttcctt cttgccacgc 3300
cccgcccaca gccccactgt gtgtgaacca agagagttgg aaggccttct gcagacggac 3360

ccattccacc tgaatctcca tgacatattg gacttctcca tataatttca tttggaaaga 3420
aacttgagca tctaaataaa atgtttggaa accaccatgc tacaagacca gcatgtgctg 3480
tctagttctt accctctgya acagtaagat tctggggcat taagacttag ttccaggatt 3540
ctgtcattct gccaacgttc tgtggctggg gttctaaagg agcttgcctg gcttagaact 3600
gcaagtgact ctagtgtgat ggagagcacc agcaaagcct tagggcccat ccctggcctc 3660
ctgttaccca cagaggggta rgcccttggc tctcttccac tatgacgtca gcttccattc 3720
ttcctttctt atagacaatt ttccatttca aggaaatcag agcccttaat agttcagtga 3780
ggtcactttg ctgagcacaa tcccataccc ttcagcctct gctccacaga gcctaagcaa 3840
aagatagaaa ctcacaactt ccttgttttg ttatctggaa attatcccag gatctggtgc 3900
ytactcagca tattcaagga aggtcttact tcattcttcc ttgattgtga ccmtgcccag 3960
sctctctgct ccctataaaa ggcaggcaga gccacygagg agcagagagg ttgagaacra 4020
cccagaaacc ttcacctctc atgctgaagc tcacaccctt gccctccaag atgaaggttt 4080
ctgcagcgct tctgtgcctg ctgctcatgg cagccacttt cagccctcag ggacttgctc 4140
agccaggtaa gacctctccc ttttaaggg gagaccaaaa gaggaattaa gaagagccat 4200
tatgtcacag ctcattagga acaaaaccag aactaaaggc tcaggtcact gaggctggtt 4260
ccccttgatct ttcctgaccc cagttttggg aggagacagt ggagccgcta cagcaacaac 4320
cctcccattg tttggggaaa taatccagaa cgaagaactg tttctcactg tgggtgtaaa 4380
ggacatttca ggccgtagtg gagagggaga aactattgcc tgaagcttca aattttggtt 4440
atggttcagt gtaccttcca gaacagtggc tgtgtaaaga ggatgaggac ccagaggaat 4500
ctcagcgtat ggcataggct aactctaaag cccatgagga tgaaagactg ggaagcaagg 4560
tattggaact tatgttccca gtgtcagaag ttttgggtta gtagacaagg actagcttgt 4620
tactcaaaat gtttccaaac ccagtcaaca atgacgggcc gcagagttca atagaggaaa 4680
gagactcaca ggcaacattt tatctctggg atctggacta agacactgaa cttgggatgg 4740
tgacttcttg gtcttctcct tccttctctt cttttcctta caaatgcaca cttacggtgg 4800
gtsctaaatg tctcattctt tgcaaaattt ctttcagatt cagtttcyat tccaatcacc 4860
tgctgcttta acgtgatcaa taggaaaatt cctatccaga ggctggagag ctacacaaga 4920
atcaccaaca tmcaatgtcc caaggaagct gtgatgtgag tggacagtgc ctggcacccc 4980
cattcaaaag ttctgatgga caacatagag aagtcaagat tcatgtccat atgagtcgga 5040
tgcatataac ttctatccaa agggcccctc taccccatag agaaactcag tccatgagaa 5100
ggagtccata actgctctag gattcccttc taggggcttg gtgaaactaa cccaatatct 5160
gtagccagga ccctggaggg tttcacctgg acagcaagag cagagcttcc ttctggagct 5220
tcttcctccc actcttcccc tccctcctct cccgggtccg ggtccttcac ctaaggacca 5280
agggctgatc agttctaggg accaatggcc cacagtcctg tgcaggatct tcaaagtctt 5340
ccatctaatt gtgccctctc tcccccacag cttcaagacc maacggggca aggagktctg 5400
tgctgacccc aaggagagat gggtcaggga ttccatgaag catctggacc aaatatttca 5460
aaatctgaag ccatgagcct tcatacatgg actgagagtc agagcttgaa gaaaagctta 5520
tttattttcc ccaacctccc ccaggtgcag tgtgacatta ttttattata acatccacaa 5580
agagattatt tttaaataat ttaaagcata atatttctta aaaagtattt aattatattt 5640
aagttgttga tgttttaact ctatctgtca tacatcctag tgaatgtaaa atgcaaaatc 5700
ctggtgatgt gtttttttgtt tttgttttcc tgtgagctca actaagttca cggcaaaatg 5760
tcattgttct ccctcctacc tgtctgtagt gttgtggggt cctcccatgg atcatcaagg 5820
tgaaacactt tggtattctt tggcaatcag tgctcctgta agtcaaatgt gtgctttgta 5880
ctgctgttgt tgaaattgat gttactgtat ataactatgg aattttgaaa aaaaatttca 5940
aaaagaaaaa aatatatata atttaaaact acttagtctt attcttcttg ggtaacatt 6000
tagctgggag tgagttttgg gcatcatggg tgacagtttg ggcatggacg ggccattttt 6060
caagaatgtc ttctggctac gctggactca accaaggttc tcagagaact tggtgggacc 6120
aggccaggat gttccagctc tctgactcta gtccctaact tcagcagccc tgattcgcta 6180
gcctctcttg tttctcttgt ttatatatta tccagcctaa ggtattttgt tatagctgcc 6240

caaaaagact aagataatct ccatcactct acccccaacc ccaatcccaa gaacttgcaa 6300
gcatccattt aaaggcgtgg aacctcttct ttttgacagc cttttaaggt caagattccc 6360
ctgtacttag tgagcttagc tgaatcttct tacaaacatg tgacccgcca tattgagcca 6420
tacataccag agcttattat ttttccagct tattgggaaa acacgtctaa ggcaaacaaa 6480
tttattgtac tgttgaacca actctgcagg gcctactgtt aatgttcatg taaaggtgcc 6540
acaattccag tatatttgga agtatagaga ggaaagagat ttgggttctg ctcctaagag 6600
taaccttaac tctcttggtc acttagggtg acataatttt tcccttcagc attaagatgc 6660
cccatccaaa aatcaggatc atactgtgtc acaggcagga agtgagagga tctagcagat 6720
gactacacaa gtccctccag ctctagagcc atgtctctta gaaacaaggt catgtttgtt 6780
gtctggggta gagtggggac ataaagttcc tgggattaaa gccataaagg cagccagggc 6840
ctagcttatt ggaaaaacac atctaagtca gataccctt atttggagac aagctttcca 6900
ttaaagttca aagtagtaca cccacagttc caagagcaat aaagcagcta agtgtcagat 6960
ggactagagg ggcacctccc aggcagggtc cacaaaaaag ctccatcaaa ctctaagctc 7020
tttcttcaat agcttcaagg tcaattgggc acagatatgt gggtcctggg cttcatgaaa 7080
tcctggacca caaatcccaa attctaaagg caagaacact cacacaaaac tgacaccagg 7140
ccacagcaaa aagacacaaa tttgttcttc tcaaatttta cccagcttca ctttgggatt 7200
atgtcattat atttcaaatg acaatgagca taatacacct taagtaacat gtcatcttat 7260
ttaagtggtt gatgttttat gtgccctgac ttgaatatat gtgttttaaa aatccacatc 7320
cttccataaa ctggttcaga tctgtaagcc ctgctttgct gtcttgtaaa ctgggggaaa 7380
gacctggaaa ggtcactcaa agctgacata caacctcagg ctcacctcag tttcaccctc 7440
agctaaggct gaccatttgt ggcagtccat gactcacttg attcactcaa tatctgcttt 7500
ctgcctcttc tcacttctca cttgctgccc ccaggcttct cctgggtgtc atcttgggtc 7560
aggagttgac tccactgcac acttgtccaa gtttcagaag tgtggaagag tgaacatttt 7620
gggtggtaac cgtccacagt gaagggtgaa ggcagatgtt caaatgcttc tgccatccct 7680
tttaccagtt ctgctctgat cttggttatt tcttttcttc tgctggggtt ggatttggtt 7740
tgttcttgtt tctctagttc cttaaggtgt gacattagat tgtctattaa tgctctttca 7800
gactttttct tttaaatata ggcatttaat gctatgagct ttcctcttag taccgttttt 7860
gctgtatccc aaaggttttg ataggttgtg tcattatcat tgttccgttc aaagatttaa 7920
aaaattttca tcttgatttc attgttgacc caacaatcat tcaggaatac attatttaat 7980
ttccatatat ttgcatggtt ttgaggtttc cttttggagt tgatttccaa ttttattcca 8040
ctttggtctg agagagtact tgatataatt tcaattttct taaatttgtt gagacttgtt 8100
ttgtggccta tcataatgtc tatcttggag aatgttccat gtgctgatga atagaatgta 8160
tactctgcag ttgttgggta gaatggtctg taaatatccc ttaagtccat ttgctctagg 8220
gtatagttta agtccattgt ttctttgttg actttctgtc ttgatgacct gtctagtgct 8280
gccattggag tattgaagtc ccccactatt attgtgttgc tatctcattt cttaggtcta 8340
gtagtaattg ttttatagat ttgggagctc cagtgttagg tgcatatata tttaggactg 8400
atattttcct gttggactag tccttttatc attatataat gtctcttttt gtctttttta 8460
attgctgttc cttcaaaatt tgttttgtct gatataaaaa tagctactcc tgctcatttt 8520
tggtgtccat ttgtatggaa tatcttcttc caccccctcta tcttaagctc atgtgagtcc 8580
taatgtgtca ggtgagtccc ctgaaaacag cagatacttg gtgaattctt atccattctg 8640
ccattctgta tttttgaagt ggagcattta ggctatttac attcaatgtt agtattgaga 8700
tgtgaggtac tactctattc ataataccat ttgttgcctg aataccttttt tttcattgtg 8760
ttattgtttt ctaagtcctg tgagatttat gctttcagga gattctagtt tggtgtattt 8820
tgtggatttg tttcaagatt tagagctcct tttagcagtt cttgtagtgc cggctcggcg 8880
gtggcaaatt ctctctgcat ttgtttgtct ggaaaagact atcttgcctt catttatgaa 8940
gcttagtttt gctggataca aaattcttgg ctgataattg ttttgtttaa gggggctaaa 9000
tatagaaccc caattccttc tagcttgtag ggttcctgct aagaaatctg ctgttaatct 9060
gataggtttt cttttatagg ttacctggta cttttgcctc acagctctta agagtctttc 9120

```
cttcctcttg actttagata atctgatgac tacttgccaa ggtgatggtc tttttacaat 9180
gaatttccca ggtgttcttt gagcttcttt tatttggatg tctgaatctc tagcaagtct 9240
gaggaagttt tcctcaaata tgttttctaa acatatttgg ggaacatatt tctaaacatt 9300
ccctcaaata tgttttctaa acttttagat ttctcttctt ccttgggaac accaataatt 9360
cttagatttg gtcgtttaac ataatctcaa gcttcttaga ggctatattc atttttttaaa 9420
ttctttttc ttcacctttg ttggattggt ttaattcaaa agccttgtct tcaag      9475
```

<210> 2
<211> 330
<212> DNA
<213> Homo sapien

<400> 2

```
atgctgaagc tcacaccctt gccctccaag atgaaggttt ctgcagcgct tctgtgcctg 60
ctgctcatgg cagccacttt cagccctcag ggacttgctc agccagattc agtttccatt 120
ccaatcacct gctgctttaa cgtgatcaat aggaaaattc ctatccagag gctggagagc 180
tacacaagaa tcaccaacat ccaatgtccc aaggaagctg tgatcttcaa gaccaaacgg 240
ggcaaggagg tctgtgctga ccccaaggag agatgggtca gggattccat gaagcatctg 300
gaccaaatat ttcaaaatct gaagccatga                                  330
```

<210> 3
<211> 109
<212> PRT
<213> Homo sapien

<400> 3

```
Met Leu Lys Leu Thr Pro Leu Pro Ser Lys Met Lys Val Ser Ala Ala
1               5                   10                  15

Leu Leu Cys Leu Leu Leu Met Ala Ala Thr Phe Ser Pro Gln Gly Leu
            20                  25                  30

Ala Gln Pro Asp Ser Val Ser Ile Pro Ile Thr Cys Cys Phe Asn Val
        35                  40                  45

Ile Asn Arg Lys Ile Pro Ile Gln Arg Leu Glu Ser Tyr Thr Arg Ile
    50                  55                  60

Thr Asn Ile Gln Cys Pro Lys Glu Ala Val Ile Phe Lys Thr Lys Arg
65                  70                  75                  80

Gly Lys Glu Val Cys Ala Asp Pro Lys Glu Arg Trp Val Arg Asp Ser
                85                  90                  95

Met Lys His Leu Asp Gln Ile Phe Gln Asn Leu Lys Pro
            100                 105
```

<210> 4
<211> 15
<212> DNA
<213> Homo sapien

<400> 4
ccctctgyaa cagta 15

<210> 5
<211> 15
<212> DNA
<213> Homo sapien

<400> 5
aggggtargc ccttg 15

<210> 6
<211> 15
<212> DNA
<213> Homo sapien

<400> 6
ctggtgcyta ctcag 15

<210> 7
<211> 15
<212> DNA
<213> Homo sapien

<400> 7
tgtgaccmtg cccag 15

<210> 8
<211> 15
<212> DNA
<213> Homo sapien

<400> 8
tgcccagsct ctctg 15

<210> 9

<211> 15
<212> DNA
<213> Homo sapien

<400> 9
gagccacyga ggagc 15

<210> 10
<211> 15
<212> DNA
<213> Homo sapien

<400> 10
tgagaacrac ccaga 15

<210> 11
<211> 15
<212> DNA
<213> Homo sapien

<400> 11
ggtgggtsct aaatg 15

<210> 12
<211> 15
<212> DNA
<213> Homo sapien

<400> 12
cagtttcyat tccaa 15

<210> 13
<211> 15
<212> DNA
<213> Homo sapien

<400> 13
ccaacatmca atgtc 15

<210> 14
<211> 15
<212> DNA
<213> Homo sapien

<400> 14
caagaccmaa cgggg 15

<210> 15
<211> 15
<212> DNA
<213> Homo sapien

<400> 15
caaggagktc tgtgc 15

<210> 16
<211> 15
<212> DNA
<213> Homo sapien

<400> 16
ttcttaccct ctgya 15

<210> 17
<211> 15
<212> DNA
<213> Homo sapien

<400> 17
gaatcttact gttrc 15

<210> 18
<211> 15
<212> DNA
<213> Homo sapien

<400> 18
ccacagaggg gtarg 15

<210> 19
<211> 15
<212> DNA
<213> Homo sapien

<400> 19
gagagccaag ggcyt 15

<210> 20
<211> 15
<212> DNA
<213> Homo sapien

<400> 20
caggatctgg tgcyt 15

<210> 21
<211> 15
<212> DNA
<213> Homo sapien

<400> 21
aatatgctga gtarg 15

<210> 22
<211> 15
<212> DNA
<213> Homo sapien

<400> 22
cttgattgtg accmt 15

<210> 23
<211> 15
<212> DNA
<213> Homo sapien

<400> 23
gagagcctgg gcakg 15

<210> 24
<211> 15
<212> DNA
<213> Homo sapien

<400> 24
tgaccatgcc cagsc 15

<210> 25

<211> 15
<212> DNA
<213> Homo sapien

<400> 25
agggagcaga gagsc 15

<210> 26
<211> 15
<212> DNA
<213> Homo sapien

<400> 26
caggcagagc cacyg 15

<210> 27
<211> 15
<212> DNA
<213> Homo sapien

<400> 27
ctctctgctc ctcrg 15

<210> 28
<211> 15
<212> DNA
<213> Homo sapien

<400> 28
agaggttgag aacra 15

<210> 29
<211> 15
<212> DNA
<213> Homo sapien

<400> 29
aaggtttctg ggtyg 15

<210> 30
<211> 15
<212> DNA
<213> Homo sapien

<400> 30
acttacggtg ggtsc 15

<210> 31
<211> 15
<212> DNA
<213> Homo sapien

<400> 31
atgagacatt tagsa 15

<210> 32
<211> 15
<212> DNA
<213> Homo sapien

<400> 32
cagattcagt ttcya 15

<210> 33
<211> 15
<212> DNA
<213> Homo sapien

<400> 33
aggtgattgg aatrg 15

<210> 34
<211> 15
<212> DNA
<213> Homo sapien

<400> 34
gaatcaccaa catmc 15

<210> 35
<211> 15
<212> DNA
<213> Homo sapien

<400> 35
ccttgggaca ttgka 15

<210> 36
<211> 15
<212> DNA
<213> Homo sapien

<400> 36
cagcttcaag accma 15

<210> 37
<211> 15
<212> DNA
<213> Homo sapien

<400> 37
tccttgcccc gttkg 15

<210> 38
<211> 15
<212> DNA
<213> Homo sapien

<400> 38
acggggcaag gagkt 15

<210> 39
<211> 15
<212> DNA
<213> Homo sapien

<400> 39
gggtcagcac agamc 15

<210> 40
<211> 10
<212> DNA
<213> Homo sapien

<400> 40
ttaccctctg 10

<210> 41

<211> 10
<212> DNA
<213> Homo sapien

<400> 41
tcttactgtt 10

<210> 42
<211> 10
<212> DNA
<213> Homo sapien

<400> 42
cagaggggta 10

<210> 43
<211> 10
<212> DNA
<213> Homo sapien

<400> 43
agccaagggc 10

<210> 44
<211> 10
<212> DNA
<213> Homo sapien

<400> 44
gatctggtgc 10

<210> 45
<211> 10
<212> DNA
<213> Homo sapien

<400> 45
atgctgagta 10

<210> 46
<211> 10
<212> DNA
<213> Homo sapien

<400> 46
gattgtgacc 10

<210> 47
<211> 10
<212> DNA
<213> Homo sapien

<400> 47
agcctgggca 10

<210> 48
<211> 10
<212> DNA
<213> Homo sapien

<400> 48
ccatgcccag 10

<210> 49
<211> 10
<212> DNA
<213> Homo sapien

<400> 49
gagcagagag 10

<210> 50
<211> 10
<212> DNA
<213> Homo sapien

<400> 50
gcagagccac 10

<210> 51
<211> 10
<212> DNA
<213> Homo sapien

<400> 51
tctgctcctc 10

<210> 52
<211> 10
<212> DNA
<213> Homo sapien

<400> 52
ggttgagaac 10

<210> 53
<211> 10
<212> DNA
<213> Homo sapien

<400> 53
gtttctgggt 10

<210> 54
<211> 10
<212> DNA
<213> Homo sapien

<400> 54
tacggtgggt 10

<210> 55
<211> 10
<212> DNA
<213> Homo sapien

<400> 55
agacatttag 10

<210> 56
<211> 10
<212> DNA
<213> Homo sapien

<400> 56
attcagtttc 10

<210> 57

<211> 10
<212> DNA
<213> Homo sapien

<400> 57
tgattggaat 10

<210> 58
<211> 10
<212> DNA
<213> Homo sapien

<400> 58
tcaccaacat 10

<210> 59
<211> 10
<212> DNA
<213> Homo sapien

<400> 59
tgggacattg 10

<210> 60
<211> 10
<212> DNA
<213> Homo sapien

<400> 60
cttcaagacc 10

<210> 61
<211> 10
<212> DNA
<213> Homo sapien

<400> 61
ttgccccgtt 10

<210> 62
<211> 10
<212> DNA
<213> Homo sapien

<400> 62
gggcaaggag 10

<210> 63
<211> 10
<212> DNA
<213> Homo sapien

<400> 63
tcagcacaga 10

<210> 64
<211> 1440
<212> DNA
<213> Homo sapien

<220>
<221> allele
<222> (30)
<223> PS1: polymorphic base C or T

<220>
<221> misc_feature
<222> (61)..(120)
<223> Ns represent sequence between polymorphic sites

<220>
<221> allele
<222> (150)
<223> PS2: polymorphic base G or A

<220>
<221> misc_feature
<222> (181)..(240)
<223> Ns represent sequence between polymorphic sites

<220>
<221> allele
<222> (270)
<223> PS3: polymorphic base T or C

<220>
<221> misc_feature
<222> (301)..(360)
<223> Ns represent sequence between polymorphic sites

<220>
<221> allele
<222> (390)
<223> PS4: polymorphic base A or C

<220>
<221> misc_feature
<222> (421)..(480)
<223> Ns represent sequence between polymorphic sites

<220>
<221> allele
<222> (510)
<223> PS5: polymorphic base G or C

<220>
<221> misc_feature
<222> (541)..(600)
<223> Ns represent sequence between polymorphic sites

<220>
<221> allele
<222> (630)
<223> PS6: polymorphic base C or T

<220>
<221> misc_feature
<222> (661)..(720)
<223> Ns represent sequence between polymorphic sites

<220>
<221> allele
<222> (750)
<223> PS7: polymorphic base A or G

<220>
<221> misc_feature
<222> (781)..(840)
<223> Ns represent sequence between polymorphic sites

<220>
<221> allele
<222> (870)
<223> PS8: polymorphic base C or G

<220>
<221> misc_feature

<222> (901)..(960)
<223> Ns represent sequence between polymorphic sites

<220>
<221> allele
<222> (990)
<223> PS9: polymorphic base C or T

<220>
<221> misc_feature
<222> (1021)..(1080)
<223> Ns represent sequence between polymorphic sites

<220>
<221> allele
<222> (1110)
<223> PS10: polymorphic base C or A

<220>
<221> misc_feature
<222> (1141)..(1200)
<223> Ns represent sequence between polymorphic sites

<220>
<221> allele
<222> (1230)
<223> PS11: polymorphic base A or C

<220>
<221> misc_feature
<222> (1261)..(1320)
<223> Ns represent sequence between polymorphic sites

<220>
<221> allele
<222> (1350)
<223> PS12: polymorphic base G or T

<220>
<221> misc_feature
<222> (1381)..(1440)
<223> Ns represent sequence 3' to PS12

<400> 64
agcatgtgct gtctagttct taccctctgy aacagtaaga ttctggggca ttaagactta 60
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 120
cctggcctcc tgttacccac agaggggtar gcccttggct ctcttccact atgacgtcag 180
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 240

```
tatctggaaa ttatcccagg atctggtgcy tactcagcat attcaaggaa ggtcttactt 300
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 360
tcttacttca ttcttccttg attgtgaccm tgcccaggct ctctgctccc tataaaaggc 420
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 480
cattcttcct tgattgtgac catgcccags ctctctgctc cctataaaag gcaggcagag 540
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 600
tgctccctat aaaaggcagg cagagccacy gaggagcaga gaggttgaga acaacccaga 660
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 720
agccaccgag gagcagagag gttgagaacr acccagaaac cttcacctct catgctgaag 780
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 840
ttccttacaa atgcacactt acggtgggts ctaaatgtct cattctttgc aaaatttctt 900
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 960
tttgcaaaat ttctttcaga ttcagtttcy attccaatca cctgctgctt taacgtgatc 1020
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1080
ctggagagct acacaagaat caccaacatm caatgtccca aggaagctgt gatgtgagtg 1140
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1200
tgccctctct cccccacagc ttcaagaccm aacggggcaa ggaggtctgt gctgacccca 1260
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1320
acagcttcaa gaccaaacgg ggcaaggagk tctgtgctga ccccaaggag agatgggtca 1380
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1440
```

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US01/29332

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) :C12Q 1/68; C07H 21/02, 21/04; C12N 15/00
US CL :435/6; 536/22.1, 23.1, 24.3; 935/76,77,78
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/6; 536/22.1, 23.1, 24.3; 935/76,77,78

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Please See Extra Sheet.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	McCARTHY et al. The use of single-nucleotide polymorphism maps o in pharmacogenetics. Nature Biotechnology. May 2000, Vol. 18, pages 505-508, see the entire document.	1-2
A	BROOKES, A.J. Review: The essence of SNPs. Gene. 1999, Vol. 234, pages 177-186, see the entire document.	1-2
A	BROOKES et al. HGBASE:A database of SNPs and other variations in and around human genes. Nucleic Acids Research. 2000, Vol. 28, No. 1, pages 356-360, see the entire document.	1-2

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search 19 NOVEMBER 2001	Date of mailing of the international search report 25 FEB 2002
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231	Authorized officer ETHAN WHISENANT, Ph.D. Dorthea Lawrence For
Facsimile No. (703) 305-3230	Telephone No. (703) 308-0196

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	VAN COILLIE et al. The human MCP-2 gene (SCYA8): Cloning, sequence analysis, tissue expression, and assignment to the CC chemokine gene contig on chromosome 17q11.2. Genomics. 1997, Vol. 40, pages 323-331, see the entire document.	1-2

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

Please See Extra Sheet.

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☒ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
1-2 (i.e. haplotypes 1 and 5)

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest ☐ The additional search fees were accompanied by the applicant's protest.
☐ No protest accompanied the payment of additional search fees.

B. FIELDS SEARCHED
Electronic data bases consulted (Name of data base and where practicable terms used):

USPATFULL, EUROPATFULL, JAPIO, MEDLINE, CAplus

SCYA8 or Small inducible cytokine subfamily , member 8 or monocyte chemotactic protein 2 AND Haplotype?

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION WAS LACKING
This ISA found multiple inventions as follows:

1. This application contains the following inventions or groups of inventions which are not so linked as to form a single inventive concept under PCT Rule 13.1. In order for all inventions to be searched, the appropriate additional search fees must be paid.

Groups I-11, Claim(s) 1-2 in part, drawn to methods for haplotyping an individual's **SCYA8** gene comprising determining which of the 11 haplotypes listed in the table in Claim 1 defines one copy of the individual's **SCYA8** gene. It is noted that Groups 1-11 correspond to the 11 haplotypes listed in the claims. For example, if Group 1 is elected, then Claims 1-2 will be examined to the extent that they apply to a method of haplotyping comprising a step of determining whether the individual has the first listed haplotype of the **SCYA8** gene. Upon selection of one or more inventions in this group, please specify the numbers of the polymorphic sites to be search.

Groups 12-23, Claim(s) 3-4 in part, drawn to methods of haplotyping the **SCYA8** gene which comprises determining which of the haplotype pairs listed in the table in Claim 3 defines both copies of an individual's **SCYA8** gene. Groups 12-23 correspond to the 12 haplotype pairs listed in the table in Claim 3. For example, if Group 12 is elected, then Claims 3-4 will be examined to the extent that they apply to methods of haplotyping comprising a step of determining whether the individual has the first listed haplotype pair. Upon election of one or more inventions in this group, please specify the haplotype pair(s) to be searched.

Groups 24-35, Claim(s) 5-7 in part, drawn to a method for genotyping the **SCYA8** gene. It is noted that Groups 25-34 correspond to polymorphic sites PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11 and PS12, respectively. For example, if Group 24 is elected, Claims 5-6 will be examined to the extent that they apply are to methods of genotyping comprising a step of identifying the nucleotide pair at PS1. Upon selection of one or more inventions in this group, please specify the numbers of the polymorphic site(s) to be search.

Groups 36-113, Claim(s) 8-9 in part, drawn to a method for haplotyping the **SCYA8** gene by identifying a **SCYA8** genotype for an individual at two or more polymorphic sites including any of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8,PS9, PS10, PS11 and PS12. It is noted that the Claims encompass methods requiring identification of possible combinations of two or more of the recited polymorphic sites, and that Groups 36-79 each correspond to one of these possible pairs, in the order recited in the Claim. For example, if Group 36 is selected, then Claims 8-9 will be examined to the extent that they apply to a combination of PS1 and PS2. Upon selection of one or more inventions in this group, please specify the site(s) to be searched.

Groups 114-191, Claim(s) 10-11, in part, drawn to a method for predicting a haplotype pair for the **SCYA8** gene by identifying a **SCYA8** genotype for an individual at two or more polymorphic sites PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9,PS10, PS11 and PS12. It is noted that the Claims encompass methods requiring identification of all possible combinations of two ormore of the recited polymorphic sites, and that Groups 114-191 each correspond to one of these possible pairs, in the order recited in the Claim. For example, if Group 114 is selected, then Claims 10-11 will be examined to the extent that they apply to a combination of PS1 and PS2. Upon selection of one or more inventions in this group, please specify the site(s) to be searched.

Group 192-215, Claim(s) 12-13 in part drawn to a method for identifying an association between a trait and one of the 11 haplotypes or one of the 12 haplotype pairs of the **SCYA8** gene. Groups 192-215 each correspond to one of the 23 particular haplotypes or haplotype pairs encompassed by the Claims For example, if Group 192 is selected, then Claims 12-13 will be examined to the extent that they apply to a method for identifying an association between a trait and the first haplotype listed. Upon selection of one or more inventions in this group, please specify the haplotype and/or haplotype pair(s) to be searched in the method for identifying an association between a trait and a haplotype and haplotype pair.

Group 216-227, Claim(s) 14-18, in part, drawn to a composition comprising at least one genotyping oligonucleotide for detecting a polymorphism in the **SCYA8** gene. If, for example, Group 216 is selected, then Claims 14-18 will be examined to the extent that they apply to an an isolated genotyping oligo for detecting a polymorphism at PS1. Upon selection of one or more inventions in this group, please specify the oligo to be searched.

Group 228, Claim 19, drawn to a kit comprising a set of oligonucleotides designed to genotype each of the polymorphic sites.

Groups 229-250, Claims 20-21 24-25 and 28, in part, drawn to a polynucleotide which is a polymorphic variant of a reference sequence for **SCYA8** gene or a fragment thereof. Claims 20-21 and 24-25 recite 11 different isogenes and 12 fragments comprising polymorphisms. If, for example, Group 229 is selected, then Claims 20-21 and 24-25 will be examined to the extent that they apply to isogene number 1. If, for example, Group 249 is selected, then Claims 20-21 and 24-25 will be examined to the extent that they apply to PS12. If, for example, Group 250 is selected, then Claims 20-21, 24-25 and 28 will be examined to the extent that they apply to the cDNA set forth in SEQ ID NO.: 2. If a member of this group is selected for examination please specify the polynucleotide to be searched.

Group 252-272, Claim(s) 22-23 and 26-27 in part, drawn to a recombinant nonhuman organism comprising a polynucleotide which is a polymorphic variant of a reference sequence for **SCYA8** gene or a fragment thereof. If, for example, Group 250 is selected, then Claims 22-23 and 26-27 will be examined to the extent that they apply to a recombinant nonhuman organism comprising isogene number 1. If, for example, Group 272 is selected, then Claims 22-23 and 26-27 will be examined to the extent that they apply to a recombinant nonhuman organism comprising PS12. If a member(s) of this group is selected for examination please specify the polynucleotide to be searched.

Group 273-275, Claim(s) 29 and 32, drawn to a polypeptide comprising an amino acid sequence which is a polymorphic variant of a reference sequence for the **SCYA8** protein or a fragment thereof. Claim 29 recites 3 distinct variant **SCYA8** proteins. If one of these groups is selected the applicant should specify the protein to be examined.

Group 276-278, Claim(s) 29, drawn to an antibody which binds to a polypeptide(s) of Claim 29. If one of these groups is selected the applicant should specify to which protein of Claim 29 the anibody binds.

Group 279-281, Claim(s) 30, drawn to a method for screening for drugs targeting the **SCYA8** polypeptides listed in claim 29. If one of these groups is selected the applicant should specify which protein of Claim 29 is being screened.

Group 282-304, Claim(s) 33, drawn to a computer system comprising polymorphism data wherein the data comprises at least one of the haplotypes and haplotype pairs listed in the claims.

Groups 305-382, claim(s) 34, in part, drawn to genome anthologies comprising two or more of the isogenes having any one of the haplotype set forth in the table. It is noted that Groups 305-382 correspond to anthologies comprising two or more of the isogenes listed in the table in claim 34 in the order shown in table. Upon selection of one or more inventions in this group, please specify the

isogene(s) to be searched.

2. The inventions listed in the instant application lack unity for a number of reasons.

a. The first claimed inventions, claims 1-2 (groups 1-10) lack unity because they represent methods which have different results depending on the nucleic acid present in the sample. That is, depending on the nucleic acid present in the sample, the special technical feature of each part of the invention would be the haplotype listed in the table. Since these methods result in different outcomes, they lack unity with one another.

b. The haplotyping methods of claims 1-2 comprises haplotyping the **SCYA8** gene to determine the presence of haplotype nuber 1. Haplotype number 1 is the reference or "wild type." Methods for "haplotyping" encompass methods in which the gene in question is sequenced, since during such a sequencing the nucleotide present at each position in the gene is determined. Since haplotype number 1 is the reference type, group I merely reads on sequencing the reference type gene. This method does not provide a special technical feature over the art, since the gene was known in the art, as were methods for sequencing the gene, as is admitted by the specification. Thus, there is no special technical feature linking the recited groups, as would be necessary to fulfill the requirement for unity of invention.

c. Each polymorphic site and each molecule containing said polymorphic site is structurally and functionally distinct from and has a different special technical feature than each other polymorphic site and molecules containing said site. The chemical structure of each polymorphism and of each molecule containing the same differ from each other. For example, a polynucleotide comprising PS1 is chemically, structurally, and functionally different from a molecule comprising PS4. As the products and methods encompassed by the claims do not share a special technical feature, the distinct products and methods may not properly be presented in the alternative. Accordingly, the claims have been separated into a number of groups corresponding to the number of different inventions encompassed by the claims, and the claims will be examined only as they read upon the invention of the elected group. For the same reasons, the remainder of the claims have been separated into a number of groups corresponding to the number of different inventions encompassed thereby.

d. The haplotypes and genotypes encompassed by the instantly recited method claims are also distinct from each other and from the single polymorphisms recited in e.g., claims 5-6. For example, a molecule of haplotype 1, comprising a particular combination of polymorphisms, differs chemically, structurally, and functionally from a molecule of haplotype 2 and from a molecule comprising a single polymorphism (e.g., PS1). The special technical feature of each haplotype or genotype is the combination of polymorphisms contained therein, which feature is lacking from and not shared with each other haplotype or genotype or with, e.g., a molecule comprising any single polymorphism set forth in the claims. Similarly, with respect to the pairs of polymorphisms of Claim 8, each combination of polymorphisms differs from each other combination and from each of the other combinations discussed above (i.e., haplotypes, genotypes, and single polymorphic sites). Thus, the claims have been separated into a number of groups corresponding to the number of different inventions encompassed thereby, and the claims will be examined only as they read upon the invention of the elected group.

e. Further, the different methods have different objectives and require different process steps. The haplotyping methods require steps of identifying haplotypes and haplotype pairs to achieve the objectives of haplotyping. The methods of genotyping require steps of identifying a single nucleotide on one gene copy to achieve the objective of genotyping. The methods of predicting a haplotype pair require steps of identifying two polymorphisms in a gene to achieve the objective of "predicting a haplotype pair". The methods of identifying an association requires steps of comparing frequencies of haplotypes in a population to achieve the objective of "identifying an association between a trait" and a haplotype. The methods of assaying for binding activity require steps of

assaying for binding activity for candidate agents. In addition to differences in objectives, effects, and method steps, it is again noted that the claims of the present Groups are not directed to the detection or identification of molecules having the same or common special technical feature, for the reasons discussed above.

f. The groups comprising polynucleotides, kits, recombinant organisms, polypeptides, antibodies, computer systems and genome anthologies are additionally drawn to multiple, distinct products lacking the same or corresponding special technical features. The nucleic acids are composed of nucleotides and function in, e.g., methods of nucleic acid hybridization or amplification. These groups are directed to different combinations of nucleic acids which are different from one another and may be employed in different methods. The recombinant organisms are complex organisms that are employed in, e.g. animal research methods. Such organisms cannot be employed as, e.g., probes or primers and they differ in both structure and function from the nucleic acids. The polypeptides differ in both structure and function from either the nucleic acids or the transgenic organisms. The polypeptides are composed of amino acids linked by peptide bonds and arranged in a complex combination of alpha helices, beta pleated sheets, hydrophobic and hydrophilic domains. The polypeptides also differ in function, e.g., fusion proteins with an enzymatic functions. The antibodies are composed of amino acids linked by peptide bonds, antibodies are glycosylated and their tertiary structure is unique, where four subunits (2 light chains and 2 heavy chains) associated via disulfide bonds into a Y-shaped symmetric dimer. The antibodies function in immunoassays. Further the computer systems are composed of, e.g., a CPU, a display device, an input device, etc., as recited in Claim 31, and function in, e.g., methods of electronic sequence comparison. Accordingly, the products differ structurally and functionally from one another. As products of different sets of Groups differ from each other in structure, function, and effect, they do not belong to a recognized class of chemical compound, or have both a "common property or activity" and a common structure as would be required to show that the inventions are "of a similar nature".